The Inclusion in a Technical Dictionary of Linguistic Information about Lexical Cohesion Phenomena in French Texts for Specific Purposes

Tanja Collet
French Studies
University of Windsor
401 Sunset
Windsor (Ontario)
Canada
N9B 3P4
tcollet@uwindsor.ca

Abstract

The purpose of this paper is twofold. Its first goal is to examine within the theoretical framework of lexical cohesion a specific behavioral pattern exhibited by French compound terms in texts for specific purposes, namely constituent deletion. Constituent deletion deletes one or more constituents of a compound term when it is used repeatedly in the same text. It reduces the standard form of the reiterated compound term to one or more shorter forms (antenne à couverture globale [earth coverage antenna] → antenne φ φ globale), which establish coreferential chains in the text. The second goal of this paper is to show how the microstructure of a monolingual technical dictionary can provide detailed information about these deletions that modify the structure of its headwords in discourse. A brief analysis of French compound terms and of constituent deletion will be carried out before the assessment of this behavioral pattern and the presentation of the microstructure that comments on it. It is our belief that the inclusion of linguistic information about the behavioral patterns of complex terms in texts for specific purposes is necessary to turn the technical dictionary into a more satisfying reference work for translators, technical writers and the like.

1. Introduction

The purpose of this paper is twofold. Its first goal is to examine within the theoretical framework of lexical cohesion a specific behavioral pattern exhibited by French compound terms in texts for specific purposes, namely constituent deletion. Its second goal is to show how the microstructure of a monolingual technical dictionary can provide detailed information about these deletions that modify the structure of its headwords in discourse. A brief analysis of French compound terms and of constituent deletion will be carried out before the assessment of this behavioral pattern and the presentation of the microstructure that comments on it.

2. French Compound Terms

French compound terms, to which we will henceforth refer by the abbreviation CT, are noun phrases or NPs belonging to the specialized lexicon. Since they are members of the lexicon, and thus words, albeit of a special nature, these NPs distinguish themselves from their truly phrasal cousins on the structural level among other things by the fact that they do not possess

an initial slot for determiners. The noun (N), which is the head of the NP, tends to take up this slot, as in the CT *guide*$_{N=head}$ *d'onde* [waveguide]. Occasionally, another lexical category, usually an adjective (ADJ) sometimes modified by an adverb (ADV), can be found in this initial position as in *haute fréquence*$_{N=head}$ [high frequency] and *très haute fréquence*$_{N=head}$ [very high frequency]. However, since French prefers head-initial NPs, N-initial CTs outnumber the other CTs, which are not N-initial. N-initial CTs also possess a larger number of increasingly complex structures made of strings, which mainly contain nouns, adjectives and prepositions (Prep). Some of these strings of varying length are N Adj, N Adj Adj, N N, N Prep N, N Prep N Adj, N Adj Prep N and N Prep N Prep N Adj, just to name a few. The CTs, which are not N-initial, on the other hand, are limited to a small number of structures, essentially Adj N, Adj N Adj and Adv Adj N. In texts for specific purposes, these lexical NPs, whether N-initial or not, are inserted into phrasal NPs, which often use determiners to introduce the lexical NPs, as can be seen in the following sentence:

[La [charge utile]$_{CT\ or\ lexical\ NP}$]$_{phrasal\ NP}$ comprend deux modules distincts, appelés respectivement module A et module B (…). (CNES/CNET 1983a, 288)

(The payload consists of two distinct modules, called module A and module B respectively (…). (literal translation proposed by the author of this paper[1]))

In other words, in the sentence above, the CT or lexical NP, *charge utile* [payload], fills the N-slot of the phrasal NP, [*la*$_{Det}$ [*charge utile*$_N$]]$_{phrasal\ NP}$.

Though they are structurally different, lexical NPs also share an important feature with their phrasal counterparts. Like phrasal NPs, they are not atomic at the level of phrasal syntax but syntactically transparent. Consequently, they exhibit low degrees of fixedness (Gross 1996), as the rules, that govern sentence and phrase formation, have retained access to their structure and its constituents. In other words, these rules do not distinguish between the phrasal NP, *la charge utile*, and the lexical NP, *charge utile*, it contains in the sentence quoted above, but treat the whole structure as a phrasal NP that can be modified.

3. Constituent Deletion

Syntactically transparent CTs behave quite distinctively in discourse. Contrary to words that exhibit syntactic atomicity and therefore a high degree of fixedness, CTs adjust their structure to the characteristics of the sentence (or sequence of sentences) in which they appear. One of the many ways in which these lexical NPs adapt to their sentential environment is by allowing the deletion of one or more of their constituents under certain circumstances. Indeed, deletions of constituents are extremely likely to occur when the sentential environment of a syntactically transparent CT presents the following conditions:

(1) it requires the reiteration of the CT; and/or
(2) it contains lexical items similar in form and/or lexical meaning to one or more constituents of the CT.

For instance, in the excerpt below, the CT *bande*$_{head}$ *de fréquences*$_{expansion}$ [frequency band] is reiterated with its head but without its expansion and takes on the form of a non-compound, namely *bande* [band]. The deletion of the expansion, *de fréquences*, seems to have been triggered by the phrasal NP *leur fréquence* [their frequency], which

immediately precedes *bande* in the sentence. The resulting structure, *ils translatent leur fréquence dans la bande 4 GHz* [they translate their frequency to the 4 GHz band], is well-formed and avoids the clumsy redondancy that would have deformed the sentence if *bande de fréquences* had been repeated as is.

> Quatre *bandes de fréquences* sont allouées au satellite Intelsat V : 14 et 6 GHz pour la réception; 11 et 4 GHz pour l'émission. (…) Après avoir amplifié les signaux en provenance des antennes, ils translatent leur <u>fréquence</u> dans la *bande ~~de fréquences~~* 4 GHz et assurent ensuite une amplification supplémentaire de ces signaux avant de les transmettre aux T.O.P. (Duigou 1977, 4-5)

> (Four *frequency bands* are assigned to the Intelsat V satellite: 14 and 6 GHz for reception; 11 and 4 GHz for transmission. (…) After amplifying the incoming signals from the antennas, they translate their <u>frequency</u> to the 4 GHz ~~frequency~~ band and then amplify the signals once more before sending them to the TWT. (literal translation proposed by the author of this paper))

4. Constituent Deletion and Lexical Cohesion

Texts, texts for specific purposes included, derive their cohesion and consequently part of their texture from the presence of coreferential chains made up of items which are identical in reference, i.e. within the text, these items refer to the same concrete or abstract thing. For their construction, these coreferential chains rely both on grammatical and on lexical means, which are often found in some kind of combination. Grammatical ways of maintaining a discourse referent are, for instance, the ellipsis of an NP or the substitution of an NP by a personal pronoun. The lexical means typically consist in the lexical repetition of the antecedent or in the use of a lexical variation, i.e. a synonym or a hyperonym, of the antecedent. A definite article or a demonstrative often precede the lexical repetition or variation and thus it can be said that their coreferentiality with the antecedent is also partly expressed through the grammar.

Lundquist (1991, 239-240) argues that in the case of "a definite NP, where the noun is *not* identical to a preceding NP, ambiguities as to whether such an expression is coreferential or not in respect to a potential antecedent may arise for the interpretation of the text." Since texts for specific purposes tend to value precision over anything else and consequently do not tolerate ambiguity, Lundquist (1991, 241) suggests "this predicts that non-ambiguous anaphoric expressions such as pronouns and definite repetitions of NPs will be the prevailing way of maintaining coreferential relations in texts for specific purposes."

Our analyses of texts for specific purposes confirm that these texts exploit the non-ambiguous grammatical and lexical means referred to by Lundquist for the construction of coreferential chains, as the two excerpts quoted below clearly show. Indeed, in the first excerpt, the pronoun *ils* [they] establishes a non-ambiguous coreferential tie with the preceding NP, *les satellites de télécommunications* [telecommunication satellites]. In the second excerpt, on the other hand, a non-ambiguous coreferential tie is created through the definite repetition of the NP, *le miroir auxiliaire* [the auxiliary mirror].

> *Les satellites de télécommunications* ont au moins un élément orienté en permanence vers la Terre et sont géostationnaires. *Ils* tournent donc naturellement, autour de l'axe Nord-Sud, à raison de un tour par jour (…). (CNES/CENT 1983b, 155)

(*Telecommunication satellites* have at least one element pointed permanently towards the Earth and are geostationary. Thus, *they* rotate naturally, about the North-South axis, making one complete rotation a day (…). (literal translation proposed by the author of this paper))

> (…) *le miroir auxiliaire* se trouve dans la zone proche de la source primaire et l'onde qu'elle rayonne est, en gros, une onde plane. Pour la transformer en onde sphérique, il faudra donc que *le miroir auxiliaire* (…). (Combes 1986, E 7 590-3)

> ((…) *the auxiliary mirror* is placed in an area close to the primary source and the wave it radiates is, approximately, a planar wave. In order to transform it into a spherical wave, it is therefore necessary that *the auxiliary mirror* be (…). (literal translation proposed by the author of this paper))

However, our analyses have also uncovered another lexical means, which neither utilizes lexical repetitions nor lexical variations (if these are to be understood as synonyms and hyperonyms), namely the substitution of a CT by another of its forms obtained through constituent deletion. For instance in the excerpt below, the CT *système de propulseurs à jet de gaz* [system of gas jet thrusters], preceded by the indefinite article *un* [a] in its phrasal NP, is shortened to *système de propulseurs à φ φ gaz* [system of gas φ thrusters] in the subsequent definite NP which is identical in reference.

> Le système AOCS utilise pour les manœuvres de régulation d'attitude (…) un volant d'inertie (…) conjointement avec *un système de propulseurs à jet de gaz* et un ensemble de senseurs terrestres (…). Pour les corrections d'orbite, c'est *le système de propulseurs à gaz* que l'on emploie (…). (Bartholomé 1979, 5)

> (The AOCS system uses for the attitude control maneuvers (…) a momentum wheel (…) together with *a system of gas jet thrusters* as well as a set of earth sensors (…). For corrections to the orbit, *the system of gas thrusters* is used (…). (literal translation proposed by the author of the paper))

This lexical way of creating cohesion conforms to the information principle which governs text production and which, similarly to the Principle of Least Effort (Zipf 1949), states that one should be brief, i.e. use the minimal amount of linguistic means sufficient to achieve one's communicative ends. But more importantly, it also avoids ambiguity as it uses different forms of the same CT to create cohesive ties. Indeed, contrary to synonyms, shortened forms are structurally similar to their antecedent as they are composed of the n number of constituents of their antecedent less the m number of deleted constituents. The structural resemblance of the two members of the cohesive tie, the proximity of the antecedent and/or the presence of lexical items with the same form and/or meaning as the

deleted constituents usually suffice to avoid any confusion. This is illustrated by the excerpt below which contains a coreferential chain made up of two synonymous CTs, one of which occurs twice, first shortened to *moteur à liquides* [liquid engine] and then in full length as *moteurs à ergols liquides* [liquid-fuel engine].

> Les raisons principales de l'adoption de ce type de propulsion [propulsion à propergol solide] en combinaison avec *des propulseurs à ergols liquides* tiennent certainement :
> - au fait qu[e] (...) les solides (...) permettent plus facilement d'atteindre des poussées élevées que *les moteurs à liquides* ;
> - au fait que les coûts de développement et les coûts récurrents de production de ces propulseurs sont généralement considérés comme inférieurs à ceux *des moteurs à ergols liquides*. (Davenas 1989, 19)

> (The main reasons behind the decision to adopt this type of rocket motor [solid-fuel motor] together with *liquid-fuel rockets* have to do with :
> - the fact that (...) solid motors (...) produce intense thrust more easily than *liquid engines* ;
> - the fact that the development costs and recurrent production costs of these motors are generally considered to be lower than those of *liquid-fuel engines*. (literal translation proposed by the author of this paper))

The coreferentiality of the shortened form, *moteur à liquides*, used 'cataphorically' in this instance, and the standard form of *moteur à ergols liquides* requires no comment, as it is obvious. The presence of synonymous CTs, albeit clearly identical in reference in this excerpt, namely *propulseur à ergols liquides* and *moteur à ergols liquides*, however does, since it seems to cast a doubt on Lundquist's hypothesis regarding the probable prevalence of non-ambiguous anaphoric expressions in texts for specific purposes. Unfortunately, we do not as yet have at our disposal statistics to confirm or infirm Lundquist's hypothesis. Nevertheless, two characteristics of French languages for special purposes (LSPs) lead us to believe that it might very well be that cohesion through constituent deletion is more widespread than cohesion through lexical variation. The two characteristics are the typical reluctance of languages for special purposes to create synonyms and the considerable number of syntactically transparent CTs, estimated at 80% by some (e.g. Boulanger 1989), LSPs contain.

To conclude this section, the analyzed excerpts indicate that a reiterated syntactically transparent CT is very likely to have its full-length form shortened by constituent deletion in French texts for specific purposes. Since lexical cohesion is partly achieved in these texts through the presence of these alternate forms, it can be said that the full-length form of a CT and its shortened forms constitute what Mortureux (1993) has termed a "designation paradigm", i.e. a paradigm of items that are identical in reference and are therefore able to construct coreferential chains in discourse.

5. Constituent Deletion and the Technical Dictionary

Technical dictionaries generally add very few linguistic comments to the conceptual information they provide. This state of affairs has been severely criticized by several terminologists, among them Slodzian (1997) and Pearson (1998), who both consider the technical dictionary "très insuffisant en matière d'information linguistique" [utterly disappointing in the linguistic information it provides], to quote Slodzian's exact words, and therefore unable to cater to the needs of its principal users: translators, technical writers, et cetera, who might be looking for advice regarding usage. In order for this reference tool to better serve its principal users, it appears necessary that it capture not only the conceptual dimension of a term but also its textual dimension. As the designation paradigm of a CT obtained through constituent deletion constitutes an integral part of the textual dimension of a term, its inclusion in dictionary entries seems essential.

We propose a basic microstructure composed of three components: headword, comment on form and comment on meaning. The lemma or headword contains the standard full-length form of a CT. The comment on form lists the shortened forms yielded by corpus analysis, which constitute non-ambiguous anaphoric or cataphoric ways of reiterating the CT in discourse. The comment on meaning, finally, provides conceptual information in the form of a definition. Cross-references to hyperonyms, hyponyms and co-hyponyms, which locate the CT within its conceptual system, could be added to the comment on meaning to facilitate access to related information.

An analysis of the following excerpt yields a designation paradigm composed of three members: *détecteur d'horizon infrarouge terrestre* [infrared earth horizon sensor], the full-length form of the CT, and two shortened forms, *détecteur infrarouge* [infrared sensor] and *détecteur d'horizon terrestre* [earth horizon sensor].

> Les antennes de télécommunications doivent être pointées vers une ou plusieurs zones qui ne sont jamais situées sur la ligne satellite-centre de la Terre. Il convient donc de maintenir une orientation déterminée autour de cet axe (...). On utilise à cet effet *des détecteurs d'horizon infra-rouge terrestre* fournissant directement les angles de tangage et de roulis. Ils mesurent l'écart de l'axe de lacet par rapport à la direction du centre de la Terre. (...) Dans ces conditions, un dépointage en lacet non détecté par *les détecteurs infrarouge* devient un dépointage en roulis un quart d'orbite plus tard, dépointage qui, lui, est détecté (...) *Les détecteurs d'horizon terrestre* sont les mieux adaptés à la précision de pointage requise actuellement pour les missions de télécommunications : ils évaluent en effet la direction du centre de la Terre et fournissent directement les erreurs de roulis et de tangage. Ils détectent la position des horizons infrarouges de la Terre (...). (CNES/CNET 1983b, 122-124)

> (Telecommunication antennas need to be pointed towards one or more areas which are never situated on the satellite-center of the Earth axis. It is therefore important that precise control be maintained about this axis (...). To that effect *infrared earth horizon sensors* are used which directly supply the roll and pitch angles. They measure the distance between the yaw axis and the direction of the Earth's center.

(...) Under these circumstances, a yaw error that went undetected by the *infrared sensors* becomes a roll error one quarter of an orbit later, error which then will be detected (...). *Earth horizon sensors* are the best adapted for the pointing precision currently required for telecommunication missions : they measure the direction of the center of the Earth and directly supply the roll and pitch errors. They detect the direction of the Earth's infrared horizons (...). (literal translation proposed by the author of this paper))

The information contained in this paragraph about the designation paradigm of *détecteur d'horizon infrarouge terrestre* is provided as follows by the microstructure:

DÉTECTEUR D'HORIZON INFRAROUGE TERRESTRE

Shortened coreferents
Détecteur d'horizon terrestre
Détecteur infrarouge

Définition
Détecteur qui évalue la direction du centre de la Terre et qui fournit les erreurs de roulis et de tangage. Il détecte la position des horizons infrarouges de la Terre et mesure dans deux plans perpendiculaires l'écart entre son axe de référence et la direction du centre infrarouge de la Terre. Au voisinage du pointage nominal, il délivre des signaux proportionnels aux dépointages en roulis et en tangage.[2]

Hyperonyms
Détecteur
Détecteur d'attitude

Co-hyponyms
Détecteur d'écart radioélectrique
Détecteur d'étoile
Détecteur solaire

Hyponyms
N/A

To complete the designation paradigm, shortened forms extracted from other sources have to be added to this preliminary entry. Below is the full designation paradigm of *détecteur d'horizon infrarouge terrestre*, as it currently appears in its entry.

Shortened coreferents
Détecteur d'horizon terrestre
Détecteur d'horizon
Détecteur infrarouge
Détecteur terrestre

The advantages of the microstructure are many. It paints a truer picture of the behavior of CTs in texts for specific purposes. By doing so, it also directly challenges terminological orthodoxy according to which "terms (...) are supposed to be labels for precise concepts within a sublanguage and, as such, not prone to variation" (Daille *et al.* 1996, 198). It

provides the user, in addition, with a more efficient reference tool: a tool that provides usage information and that, if it comes with an index, allows the user to access the information he or she requires from any of the forms a CT might have in discourse.

To conclude this section, here is the preliminary entry of *antenne à couverture globale* [earth coverage antenna], which also contains a designation paradigm, albeit much smaller in size than the paradigm associated to *détecteur d'horizon infrarouge terrestre.*

ANTENNE À COUVERTURE GLOBALE

Shortened coreferents
Antenne globale

Definition
Antenne de satellite de télécommunications dont le faisceau couvre la totalité du globe terrestre. Pour un satellite géostationnaire de télécommunications, ceci correspond à une ouverture angulaire de faisceau de 17,34°.[3]

Hyperonyms
Antenne
Antenne de télécommunications
Antenne de satellite de télécommunications
Antenne directive

Co-hyponyms
Antenne pour couverture d'hémisphères et de zones

Hyponyms
N/A

6. Conclusion

The basic microstructure proposed in this paper is an attempt to provide the dictionary user with information about at least one aspect of the textual dimension of CTs, namely the cohesive value of constituent deletion. It stems from the belief that the inclusion of such linguistic information is necessary to turn the technical dictionary into a more satisfying reference work for translators, technical writers and the like.

Endnotes

1 The literal translations proposed by the author for the benefit of readers less familiar with French are "rough translations". They are, by no means, to be regarded as "finished translations" of the original French excerpts.
2 The definition in this preliminary version of the entry is based on CNES/CNET (1983b, 124).
3 The definition in this preliminary version of the entry is based on Nouvel (1983, 56).

References

Bartholomé, P. 1979. « Le programme européen OTS de satellite de télécommunications. Développement du satellite et performances en orbite », *L'Aéronautique et l'Astronautique*, 2, 3-14.

Boulanger, J.-C. 1989. « Le statut du syntagme dans les dictionnaires généraux monolingues », *Meta*, 34 (3), 360-369.

CNES/CNET 1983a. *Télécommunications spatiales III. Secteur terrien. Systèmes de télécommunications par satellite*. Paris : Masson.

CNES/CNET 1983b. *Télécommunications spatiales II. Secteur spatial*. Paris : Masson.

Collet, T. 2003. 'A Two-Level Grammar of the Reduction Processes of French Complex Terms in Discourse', *Terminology*, 9 (1), 1-27.

Combes, P.-F. 1986. « Antennes pour techniques spatiales », *Techniques de l'Ingénieur, traité Électronique, E7 II Electronique-Télécommunications*, 590-1-590-21.

Daille, B. *et al.* 1996. 'Empirical Observation of Term Variations and Principles for their Description', *Terminology*, 3 (2), 197-257.

Davenas, A. 1989. « La technologie des propergols solides et ses applications spatiales », *L'Aéronautique et l'Astronautique*, 5, 19-26.

Duigou, M. 1977. « Le satellite Intelsat V », *L'Aéronautique et l'Astronautique*, 1, 3-9.

Gross, G. 1996. *Les expressions figées en français. Noms composés et autres locutions*. Paris : Éditions Ophrys.

Lundquist, L. 1991. 'Some Considerations on the Relations between Text Linguistics and the Study of Texts for Specific Purposes' in H. Schröder (ed.), *Subject-oriented Texts. Languages for Special Purposes and Text Theory*. Berlin/New York: Walter de Gruyter, 231-243.

Mortureux, M.-F. 1993. « Paradigmes désignationnels » in S. Moirand et J. Peytard (eds.), *Configurations discursives, Semen 8*. Paris : Université de Besançon/Les Belles Lettres, 121-141.

Nouvel, M. 1983. *Dictionnaire bilingue des télécommunications internationales. 1. Théorie des antennes*. Montréal : Teleglobe Canada.

Pearson, J. 1998. *Terms in Context*. Amsterdam/Philadelphia: John Benjamins.

Slodzian, M. 1997. « Un obstacle à la conception de nouveaux outils: l'approche discontinue et fragmentaire de l'objet 'dictionnaire' » in A. Hermans (ed.), *Les dictionnaires spécialisés et l'Analyse de la Valeur*. Louvain-La-Neuve : Peeters, 23-31.

Zipf, G. K. 1965. *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*. (facsimile of 1949 edition) New York/London: Hafner Publishing Company).