UDC 81'373.74
811.111.374
811.133.374
Original scientific paper
Received on 20.10. 2005.
Accepted for publication 30.11. 2005.

Mojca Pecman
Centre National de Recherche Scientifique Bases,
Corpus et Langage
Université de Nice-Sophia Antipolis

Systemizing the notation and the annotation of collocations

The present article is an attempt at systemizing the notation and the annotation of collocations. In spite of the ever increasing number of works devoted to lexicology, lexicography and, more specifically, phraseology, the question of markers allowing the extraction of units from their phrasal context and of labels destined to specify the usage of units in the context of a language is rarely addressed. The authors and the designers of dictionaries usually develop their own set of markers and labels in order to satisfy the publishing needs but their use remains seldom systematized. The present article offers a critical examination of the use of markers and labels in lexicography and presents the solutions adopted on the matter within a research project carried out at the University of Nice (Pecman 2004a). The conclusion will put forth the importance of the design of systematic and rational modelling procedures for the processing of collocational resources.

Key words: phraseology; collocations; lexicography, notation of collocations; annotation of collocations; processing of collocations; modelling of collocations; dictionary making; Foreign Language Teaching; French language; English language.

1. Introduction

The use of markers for the extraction of phraseological units (PUs) from their phrasal context (e.g. *sth, sb, sb's* in English*, qch, qn* in French, etc.) and of labels destined to specify the usage of units in the context of a language (e.g. *démentir une hypothèse* <fonction: predicate>, <type of discourse: academic and

scientific>, <hyperonym: hypothesis>, <synonym: *contredire une hypothèse*>, <antonym: *confirmer/conforter une hypothèse* >, <English equivalent: *to invalidate a hypothesis*>, etc.) is one of the key issues of modern lexicography, whether we consider monolingual or bilingual lexicography. The renewed interest for the phraseological approach of languages and the recent appearance of collocational dictionaries (cf. Benson *et al.* 1997, Cowie *et al.* 1983, Cowie & Mackin 1975), also called combinatorial dictionaries (cf. Mel'čuk *et al.* 1999, Zinglé 2003), fostered discussion on this crucial question.

The use of markers and labels within the framework of lexicographical processing of collocations is related to the process of modelling of lexical resources (cf. Pecman 2005). In that instance, the markers and the labels comprise two closely related but distinct stages in the range of subsequent steps which form the complex procedure of collocational resources processing, notably the extraction, the notation, the annotation, the presentation and the exploitation of collocations.

In spite of the ever increasing number of works devoted to lexicology, lexicography and, more specifically, phraseology, the notation and the annotation of collocations remain stumbling blocks in our common effort to create re-exploitable collocational resources.

In order to draw attention to various problems in relation with the marking and labelling of collocational resources, the article presents, first, a general methodological framework, then a critical analysis of the use of markers and labels in lexicography, and ultimately the solutions adopted on the matter during a research project carried out at the University of Nice (cf. Pecman 2004a), with special emphasis on the advantages of semantic labelling of collocations.

2. Methodological background

The role of marking and labelling of collocational resources is examined within the framework of an empirical study of English-French phraseology for academic and scientific purposes. The goal of such an application setting is to offer French academics and scientists a tool for an easy access to English routine formulas in that specific genre (Pecman 2004a & 2004b).

In order to be able to investigate the general scientific phraseology from the contrastive point of view, we have designed a parallel corpus containing 82,800 words. The textual sources are scientific in nature (cf. scientific articles, abstracts, activities reports, communications…) and are taken from three related

domains: physics, chemistry and biology. The corpus was set up in order to allow the observation of phraseological properties of the English for Academic Purposes (EAP) and English for Science and Technology (EST). In the context of the English language, there are few linguists working on the phraseology of this sub-language, namely Granger (1998) and Howarth (1996). In France, this specific genre analysis gave rise to the study of what Phal (1971) had called "Vocabulaire Général d'Orientation Scientifique" (V.G.O.S.). We refer to this sublanguage by the term General Scientific Language (GSL) (cf. Pecman 2004a: 124).

This corpus was then used for the extraction of bilingual phraseological units (i.e. units of translation) with no restriction on whether the equivalences within the unit behave as trivial lexical correspondences or as lexical mismatches and with no restriction on whether the units play the role of a predicate, a noun, an adjective, an adverb, a quantifier, a preposition or a conjunction (e.g. *this work is an outgrowth of → ce travail découle de, to develop a new method → [mettre au point/développer] une nouvelle méthode, crucial feature → caractéristique principale, familiar phenomenon → phénomène courant, time-consuming → exigeant beaucoup de temps, fairly complicated → assez compliqué, in the long-term → à long terme, in close collaboration → en étroite collaboration, a wide range of → une vaste gamme de, with respect to → en ce qui concerne, par rapport à; at the same time → en même temps, as expected → comme prévu...*). The major criteria on which the extraction of multiword units was based are their frequency, their diffusion in each of the three domains and their re-exploitability during the writing process.

The corpus was retrieved both manually and automatically. The machine processing was completed with the software ZText (Zinglé, 1998). The results obtained from these two compilation procedures were confronted and organised in the form of a bilingual phraseological database. The lexical resources were then verified and corrected with the help of a number of monolingual classical and collocational dictionaries (cf. *The BBI Dictionary of English Word Combinations* by Benson *et al.* 1997, *Oxford Collocations Dictionary for students of English* 1997, *Selected English Collocations* by Hill & Lewis 2002, *Le Trésor de la Langue Française Informatisé* 2002). At this stage of corpus processing, the size of the database had reached the number of some 2000 units of translation. Nevertheless, this number is still growing as we are continuing to update our data currently. In order to increase the exploiting potential of our resources, we have recently embarked on an additional collection of data from comparable corpora. This supplementary technique should allow to compensate for the insufficiencies that parallel corpora suffer, namely the high number of mistranslations, paraphrasing, etc.

The extraction of PUs and the creation of a bilingual phraseological database lead to a design of a model for systemizing the notation and the annotation of data. Previous to the design itself was a reflection on the role and the use of markers and labels in bilingual lexicography.

3. Markers of contextual anchoring

The present section offers a critical analysis of the abbreviations (e.g. *sth, sb, sb's* in English*, qch, qn* in French, etc.) and other means of marking (such as ellipsis) used in bilingual dictionaries for the purposes of the extraction of lexis from their phrasal context. The analysis is based on two bilingual French-English/English-French dictionaries, namely Harrap's dictionary (1997) and Oxford-Hachette dictionary (1994-1996).

Our observations are based on a list of abbreviations which act as contextual anchors (in contrast with grammatical and categorial abbreviations such as *v, vt, n, adj,* etc. which are used for specifying the nature of a word indexed in the dictionary: e.g. *lunch* ***n***) and which were found in the articles of the two dictionaries. At the time of actualization of an expression, these abbreviations yield their place to an unspecified element of the language, provided that the latter one is semantically compatible with the rest of the expression and of the same nature as the marker in question (thus for example *to take* ***sth*** *to bits* allows to construct expressions such as *to take* ***a car*** *to bits, to take* ***a toy*** *to bits,* etc. but not **to take* ***interesting*** *to bits,* **to take* ***Mary*** *to bits* or **to take* ***a cake*** *to bits*). One could refer to these abbreviations in terms of "variables", insofar as they serve for defining a "procedure" without knowing beforehand the data which will take the place of the variable during the execution of the procedure. From the grammatical point of view, the elements which can be replaced by a marker are various: we find, for example, verbs' objects, animated or inanimate (e.g. *to praise* ***sb****, to return* ***sb's*** *call, to leave* ***sth*** *unfinished*), sets formed by a verb and its object playing a role of a support to an adverbial (e.g. ***to do sth*** *for a laugh*) or whole propositions (e.g. *to acknowledge* ***that…***). If we refer, however, to the list of abbreviations given at the beginning of the two dictionaries, there are very few of them mentioned: *sth, qch* and *qn* are the only abbreviations mentioned in Oxford-Hachette and *sb, sth, qch* and *qn* are the only abbreviations mentioned in Harrap's. And yet, more than ten different markers can be found in each of the two dictionaries (see Tables 1 and 2).[1]

[1] Regarding *one's* and *oneself* (e.g. *to try* ***one's*** *luck* → *tenter* ***sa*** *chance, to protect* ***oneself*** → ***se*** *protéger*), it is important to note that their role is similar to that of abbreviations in the entries of dictionaries, insofar as, due to their pronominal nature, *one's* and *oneself* can easily be

The comparison of dictionaries reveals many disparities in the use of markers of contextual anchoring. Firstly, the two dictionaries share a certain number of markers: *sth, sb, sb's, to do sth* and *doing sth*. Secondly, some markers appear in different forms: *that…* and *that, to do sth* and *to do, doing sth* and *doing, sb/sth* and *[sb/sth], of sth/that* and *of/that, 'at doing sth* or *to do sth'* and *'to do* ou *at doing'*. Thirdly, certain markers exist only in one of the two dictionaries: *to (ou for) sth/to do sth, [sth], [sb], on sth/on doing.*

Furthermore, we can note certain tendencies in the use of the markers in each of the dictionaries, in particular regarding the expression of object. For example, Harrap's has a tendency to state the object: *to do sth, doing sth*, *of sth/that…, 'at doing sth* or *to do sth'*, contrary to Oxford-Hachette who often avoids it: *to do, doing, of/that, 'to do* ou *at doing'*. As a matter of fact, in the latter one, the presentation is less homogeneous since the object is sometimes noted, sometimes omitted: one finds for example at the same time *to do* and *to do sth, doing* and *doing sth*. Similarly, the two dictionaries use alternatively the slash and the conjunction *or* to note the possibility of a choice between two formulas: in Harrap's we find *to sth/to do sth* and *'at doing sth* or *to do sth'* and Oxford-Hachette proposes *on sth/on doing* and *'to do* ou *at doing'* (the latter one adopts the point of view of the user by expressing the conjunction in French).

Finally, not only are the majority of the abbreviations not listed on the preliminary pages, but in neither of the two dictionaries are they distinguished by a specific typography, in much the same way as no explanation is provided as to their use. It would however be helpful to specify, in the case of the Oxford-Hachette in particular, the exact meaning of [sth], [sb], [sb/sth] that one finds in examples such as *break [sth] down, break down [sth]*; *break [sb] in* and *set [sb/sth] down*.

---

replaced by another element of the language (e.g. *to try* ***one's*** *luck* → *he tried* ***his*** *luck, to protect* ***oneself*** → *he* ***protected*** *himself*), in the very same way as the French possessive and reflexive pronoun *se* can be replaced by another possessive or reflexive pronoun (*tenter* ***sa*** *chance* → *j'ai tenté* ***ma*** *chance,* ***se*** *protéger* → *je* ***me*** *suis protégé*). However, due to the time constraints, we will not examine the use of these pronouns in the present study. We will just draw attention to the fact that there are various elements in the language which adopt certain, more or less, canonical form within an expression listed in a dictionary and which are likely to be modified (through conjugation, agreement, etc.) at the moment of actualisation (such as, already mentioned pronouns, infinitives, first person singular, indefinite articles, etc.)

HARRAP'S DICTIONARY		
ABBREVIATION	**EXAMPLE IN ENGLISH**	**PROPOSED TRANSLATION IN FRENCH**
sth	*to leave **sth** unfinished* *to take **sth** to bits*	*laisser **qch** inachevé* *démonter **qch***
sb	*to praise **sb** lavishly* *to take **sb** by the hand/arm*	*couvrir **qn** d'éloges* *prendre **qn** par la main/bras*
sb's	*to sit on **sb's** lap* *to break **sb's** heart*	*s'asseoir sur les genoux **de qn*** *briser le cœur à **qn***
that…[2]	*it is clear **that…*** *to take the attitude **that…***	*il est clair* ou *évident **que…*** *considérer **que…***
to do sth	***to do sth** for a laugh* *to decline **to do sth***	***faire qch** pour rire* *refuser de **faire qch***
doing sth	*to be good/bad at **doing sth*** *to take pleasure in **doing sth***	*être/ne pas être doué pour **faire qch**, prendre plaisir à **faire qch***
sb/sth	*to lean over **sb/sth*** *to give **sb/sth** a shake*	*se pencher par-dessus **qn/qch*** *secouer **qn/qch***
to sth/to do sth[#3]	*to consent **to sth/to do sth***	*consentir **à qch/à faire qch***
for sth/to do sth[#]	*to be set **for sth/to do sth***	*être prêt **pour qch/à faire qch***
of sth/that…	*to be conscious **of sth/that…***	*être conscient **de qch/que…***
at doing sth or *to do sth*	*to make an attempt **at doing sth** **or** **to do sth***	*essayer* ou *tâcher **de faire qch***[4]

Table 1. The use of markers (*sth, sb, sb's, qch, qn,...)* in Harrap's dictionary

[2] The case of *that...* is particular insofar as *that* is not replacing another element of the language but announces the obligatory addition of another element of the language, namely of a proposition. In fact, it is the suspension points that play the role of "variable."

[3] Symbol # indicates that this abbreviation exists only in one of the two dictionaries, Oxford-Hachette or Harrap's.

[4] Curiously, this example is set up to give the impression that the choice between grammatical forms in English (infinitive or -ing form) is parallel to the choice between lexical items in French (*to test* or *try*).

OXFORD-HACHETTE DICTIONARY		
ABBREVIATION	EXAMPLE IN ENGLISH	PROPOSED TRANSLATION IN FRENCH
sth	*to make a study of* ***sth***	*faire une étude de* ***qch***
[sth] [#]	*break* ***[sth]*** *down, break down* ***[sth]***	*enfoncer [door]; démolir [fence, wall], etc.*[5]
sb	*to talk* ***sb*** *into doing*	*persuader* ***qn*** *de faire*
[sb] [#]	*break* ***[sb]*** *in*	*accoutumer* ***[qn]*** *au travail*
sb's	*to return* ***sb's*** *call*	*rappeler* ***qn***
that[##6]	*it's difficult to accept* ***that***	*on a du mal à accepter* ***que (+ subj)***
to do sth	***to do sth*** *as a joke*	***faire qch*** *par plaisanterie*
to do[##]	*to find it difficult* ***to do*** *to consent* ***to do***	*avoir du mal* ***à faire*** *consentir* ***à faire***
doing[##]	*it's no joke* ***doing*** *pleasure of* ***doing***	*ce n'est pas drôle* ***de faire*** *plaisir* ***de faire***
doing sth	*to consent to sb* ***doing sth***	*consentir à ce que qn* ***fasse qch***
on sth/on doing[#]	*to be set* ***on sth/on doing***	*tenir absolument* ***à qch/à faire***
[sb/sth] [##]	*set* ***[sb/sth]*** *down*	*déposer [passenger]; poser [suitcases, vase]*
of/that[##]	*to make sb aware* ***of/that***	*rendre qn conscient* ***de/que***
to do ou *at doing*[##]	*to make an attempt* ***to do*** **ou** ***at doing***	*tenter* ***de faire***

Table 2. The use of markers (*sth, sb, sb's, qch, qn,...)* in Oxford-Hachette Dictionary

4. Set of markers designed *ad hoc*

The creation of a bilingual phraseological database (focused on English and French) destined to generate tools for assisted scientific writing (Pecman 2004a), has lead us to consider the problem of the use of markers of contextual anchoring. Due to the general absence of a uniform system of marking which would be recognized as a national, or even international, standard on behalf of the lexicolographers, we were obliged to develop our own set of markers to pursue the process of data collection. In phraseology, an appropriate use of markers

[5] In Oxford-Hachette, the specification on the contexts in which a word can be used are stated in the source language.

[6] The symbol ## indicates that this abbreviation exists in Harrap's in another form.

is all the more important as the phraseological resources are, by definition, multiword units: their collection is based essentially on our capacity to extract this type of knowledge from textual corpora and to model it in order to ensure its exploitability (e.g. *We take so many things for granted nowadays.* thus allows, with the help of the marker of anchoring *sth* to extract the phraseological unit *to take sth for granted*).

A preliminary design of a set of markers was carried out with the aim to harmonize their use in the database. In order to distinguish them from the rest of the phrase, all the markers were put in angle brackets (Table 3 illustrates the main ones.)

CODE		MEANING	EXAMPLE
ENGLISH	FRANCH		
<sth>	<qch>	something quelque chose *ou* de faire quelque chose	*to prove useful for <sth>* → *s'avérer utile pour <qch>*
	<faire qch>		*to strive for <sth>* → *s'efforcer de <faire qch>*
<doing sth>	<qch>	doing something quelque chose *ou* de faire quelque chose	*to cause discrepancies in <doing sth>* → *causer des écarts dans <qch>*
	<faire qch>		*towards the goal of <doing sth>* → *dans le but de <faire qch>*
<sb>	<qn>	somebody quelqu'un	*to give <sb> an in-depth understanding of <sth>* → *permettre à <qn> de comprendre en profondeur <qch>*
<sb's>	de <qn>	somebody's de quelqu'un	*to apply <sb's> law* → *appliquer la règle de <qn>*
to <do sth>	de/à/pour <faire qch>	to do something de/à/pour faire quelque chose	*to have the potential <to do sth>* → *avoir la possibilité de <faire qch>*
			our strategy is to <do sth> → *notre stratégie consiste à <faire qch>*
			to be used <to do sth> → *être utilisé pour <faire qch>*

Table 3. The markers developed *ad hoc*

This model accounts for the harmonization of the use of markers: all the elements which are replaced at the time of the communication appear between angle brackets while the stable elements appear outside the brackets, optional markers are separated by the slash (and eventual non compulsory elements could appear between simple brackets). Moreover, this is an extensible model: if need be, one can easily add more markers which conform to the construction pattern.

5. Labels

A design of labels is another essential step in the process of constructing re-exploitable lexical resources. Any creation of re-exploitable lexical database, whether we chose to store simple lexical items or complex ones, is necessarily subjected to a preliminary reflection on the types of labels to adopt. The result of this reflection, i.e. the scope of the selected labels and the method for their assignment to the lexis, are related to scientific objectives which underline a lexicographical project.

Thus, within the framework of the research undertaken at "Centre de Formation des Traducteurs et Terminologues" of the University of Rennes II, the team of Daniel Gouadec, who has a rich experience in the processing of scientific and technical terminology, sought to align the modelling of the PUs to the modelling of terms. In this respect, it was important to place at the disposal of the users the same quantity and the same quality of information for the two types of data, terms and PUs. Consequently, the labels chosen for the modelling of the latter ones are extremely varied (cf. Gouadec, 1993: 178-189) and they tend to take into account the totality of information contained or relative to such and such unit: the model consists of the labels of usage, of source, of nature, of function, of concept, of meaning, of register, of connotation, etc, as well as of cross-references to correlative units, such as synonymic PU, identical PU, antonymic PU, generic PU and specific PU. These labels are considered by the team of Gouadec as potential labels, since for certain lexical items, certain labels are irrelevant.

6. Set of labels designed *ad hoc*

We have first examined the advantages and the disadvantages of a broad labeling system. In the view of our specific scientific objective—namely the validation of a processing method involving collocations for the purposes of foreign

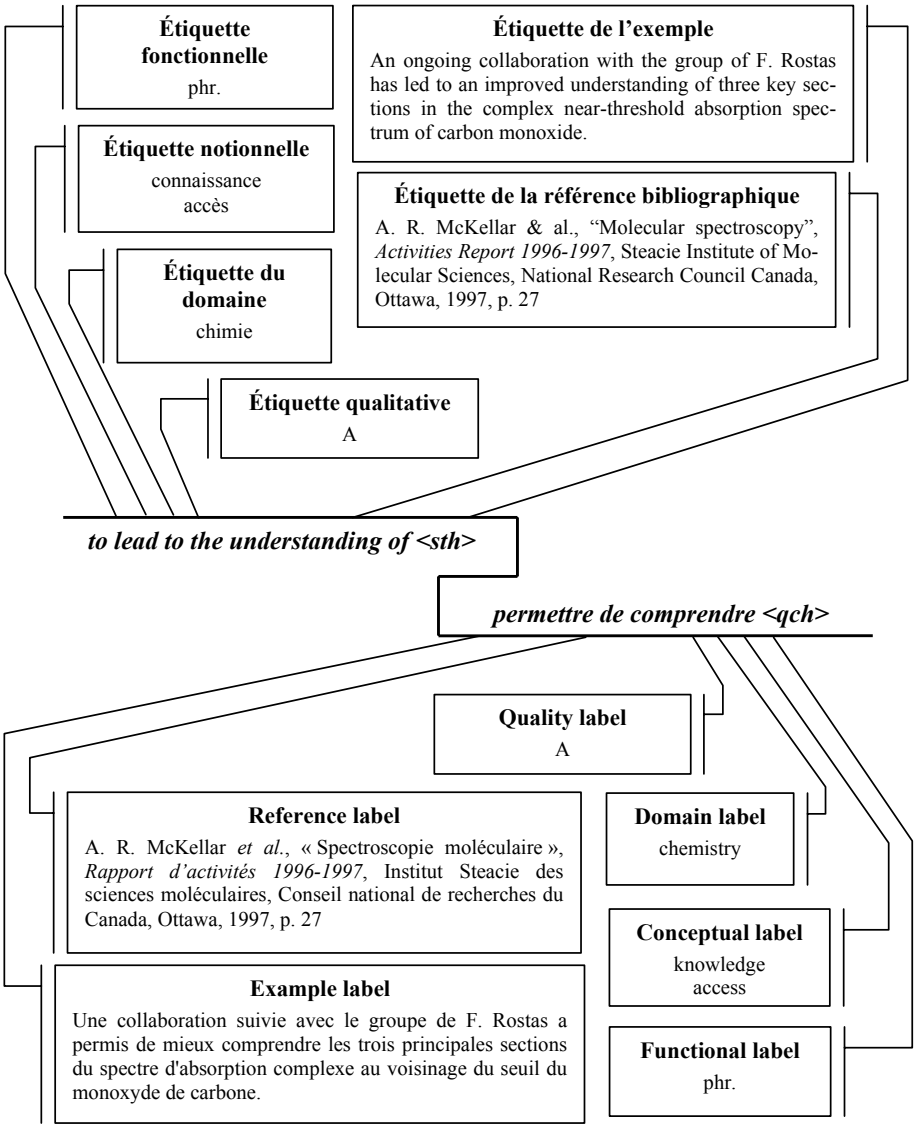

Figure 1. Labelling of collocational resources

languages learning and writing (Pecman 2004b)—we have decided to retain only those labels which were relevant for the entirety of our data, insofar as this data could be used for the creation of a phraseological dictionary of general scientific language (cf. Pecman 2004a, 2004b). Consequently, a set of six labels was designed (cf. Figure 1) which accounts for the formalisation of PUs. Each

of the six labels is associated to every unit of translation providing information, namely on the syntactic function of the unit in the discourse (functional label), on the meaning of the unit (conceptual label), on the adequacy of the unit as to the status of the unit of translation (quality label), on the scientific field the source text belongs to (domain label), on the sentential context from which the units were extracted (example label) and on the bibliographical reference of the source text (reference label). The first three labels imply a classification of collected units: the functional label, the conceptual label and the quality label exploit the inherent properties of PUs; on the contrary, the last three labels: the domain label, the example label and the reference label, refer to extralinguistic information in relation with PUs. (N.B. The names of the labels associated to French PUs are expressed in English and vice versa.)

PUs are stored in electronic form and coded with the formalism which allows their exploitation with softwares ZDoc (Zinglé, 1999) and ZLoc (Zinglé, 1998) which are part of Z-station workbench (Zinglé, 1944). The function of labels is twofold: on the one hand, they simplify the analysis of the data and, on the other hand, under the circumstances of an exploitation of resources for the creation of a bilingual collocational dictionary, they can offer to eventual users of the dictionary helpful information on the usage of PUs.

The main advantage of the proposed model is the possibility of accessing data on semantic bases, thanks to conceptual labelling.

7. Semantic labelling of collocations

The semantic categorisation of multiword units is a part of a project aiming at the construction of a dictionary that offers a flexible approach to resources. Such a dictionary should offer, besides the classical alphabetic access to data, the possibility of accessing lexical resources through semantic query.

This claim was taken into account during the process of formalisation of resources as every multiword unit was labelled semantically. Altogether, 125 semantic categories have been identified so far. In order to give an overview of the conceptual organisation of general scientific discourse, we have constructed an ontology (cf. Pecman 2004a: 297-303) devoted to this specific "discourse community" (cf. Swales 1990).

As a matter of fact, the model is based on the semantic component of the language and consists in linking every multiword unit to a conceptual condensed representation of its dominant meaning, more precisely to its hyperonymic

synonym. The aim of the model is to offer potential users a flexible approach to collocations: one semasiological, allowing them to access data from their form and one onomasiological, providing an access key to the same data from their meaning. In the former case, a multiword unit like *it is widely accepted that* can thus be located through its lexical constituents *widely* and *accept* and consulted together with other multiword units with which it shares one of them (i.e. *to accept sth fully/readily*, *to accept a criterion/condition*, *to accept a transformation/modification*, etc.). In the latter case, the same multiword unit can also be accessed through its hyperonymic synonym, in this instance coded as |QUOTATION|, and found in an entry together with other units of similar meaning (i.e. *it is commonly/generally/universally/widely accepted that*, *it is widely/well known that*, *it [is/has been] (often) asserted/noted/recognised/believed/claimed/argued that*, etc.). Each of these access points is meant to offer a pathway to French equivalents and vice versa, in this particular case to or starting from the following units: *il est commun de penser que, il est communément/généralement/unanimement admis que, on admet que, on a longtemps pensé/cru que, on a souvent dit que*, etc.

8. Conclusion

Our study on the use of markers of contextual anchoring and labels for specifying the usage of units within a language hopes to draw the attention of lexicographers to the urgent need for systematizing the notation and the annotation of lexical resources.

Our critical analysis of the two well known English-French/French English dictionaries, namely Oxford-Hachette and Harrap's, points to the important lack in homogenization of the way resources are coded: the markers for the extraction of phraseological units from their phrasal context (e.g. *sth, sb, sb's* in English*, qch, qn* in French, etc.) which are employed, reveal a total absence of a preliminary reflection on the question. In parallel, the labelling of phraseological units is often a matter of independent ventures. Yet, an all embracing and widely accepted system of notation and annotation accounts for the homogeneity within lexical resources and thus guarantees their re-exploitability, whether the resources are designed for the purposes of linguistic analysis or for the purposes of second language teaching tools' design. The model we have illustrated in this paper shows an attempt at systemizing the notation and the annotation of collocational resources and provides the main guidelines which can be followed for designing systematic and rational modelling procedures for collocational resources.

The progress in phraseology and the success of many projects aiming at the construction of re-exploitable collocational resources depend on our capacity to solve the difficult question of the use of markers and labels.

References

Benson Morton, Evelyn Benson & Robert Ilson (1997). *The BBI Dictionary of English Word Combinations*. 2nd ed. [1986 for the 1st ed.]. Amsterdam - Philadelphia: John Benjamins.

Cowie Anthony P. & Ronald Mackin (1975). *Oxford Dictionary of Current Idiomatic English*. Volume 1. 2nd ed. London: Oxford University Press.

Cowie Anthony P., Ronald Mackin, Isabel R.I. McCaig (1983). *Oxford Dictionary of Current Idiomatic English*. Volume 2. London: Oxford University Press.

Gouadec, Daniel (1993). Extraction, description, gestion et exploitation des entités phraséologiques. *Terminologies Nouvelles*, n. 10, Actes du séminaire international. (Hull, mai 1993). Rint, 16-22.

Granger, Sylviane (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. Cowie Anthony P., ed. *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, 145-160.

Harrap's Compact Dictionary: anglais-français/français-anglais (1997). Edinburgh: Chambers Harrap Publisher Ltd.

Hill, Jimmie, Michael Lewis, eds. (1997). *Dictionary of Selected Collocations*, Based on the original work of C.D. Kozlowska and H. Dzierżanovska, Hove: Language Teaching Publications.

Howarth, Anthony P. (1996). *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Tübingen: Niemeyer.

Le Trésor de la Langue Française Informatisé (2002). Version 4 du 10 décembre 2002, Analyse et Traitement Informatique de la Langue Française (A.T.I.L.F.), C.N.R.S. (http://atilf.atilf.fr).

Mel'čuk, Igor, Nadia Arbatchewsky-Jumarie, Lidija Iordanskaja, Suzanne Mantha, Alain Polguère (1999). *Dictionnaire explicatif et combinatoire du français contemporain: recherches lexico-sémantiques IV*, Montréal : Les Presses de l'Université de Montréal.

Oxford Collocations Dictionary for Students of English (2002). Oxford: Oxford University Press.

Oxford-Hachette Dictionary (1994-1996). Français-Anglais, Anglais-Français, version 1.1, Oxford University Press - Hachette Livre (version électronique).

Pecman, Mojca (2004a). *Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique*, Thèse de doctorat, 9 déc. 2004, Dir. Henri Zinglé, Université de Nice-Sophia Antipolis.

Pecman, Mojca (2004b). Exploitation de la phraséologie scientifique pour les besoins de l'apprentissage des langues. *Actes de la Journée d'étude de l'ATALA. Traitement Auto-*

matique des Langues et Apprentissage des Langues. 22 octobre. Université de Grenoble Stendhal, 145-154.

Pecman, Mojca (2005). Compilation, formalisation and presentation of bilingual phraseology: problems and possible solutions. *The Many faces of Phraseology. An Interdisciplinary Conference*, 13-15 October, Louvain-la-Neuve, Belgium, Amsterdam - Philadelphia: John Benjamins (forthcoming).

Phal, André (1971). *Vocabulaire général d'orientation scientifique (V.G.O.S.) : part du lexique commun dans l'expression scientifique*. Paris: CREDIF.

Swales, John M. (1990) *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Zinglé Henri (2003). *Dictionnaire combinatoire du français: Expressions, locutions et constructions.* Paris: La Maison du Dictionnaire.

Zinglé, Henri (1994). The Z-station workbench and the modelling of linguistic knowledge. Carlos Martin Vide, ed. *Current issues in mathematical linguistics*. Amsterdam: North-Holland, 423-432.

Zinglé, Henri (1998). ZTEXT: un outil pour l'analyse de corpus. *Travaux du LILLA*, n. 3, Publications de la Faculté des Lettres, Arts et Sciences Humaines de l'Université de Nice-Sophia Antipolis, 69-78.

Author's address:

UMR 6039 Bases, Corpus et Langage
Université de Nice
98 bd Edouard Herriot - BP 320
06204 Nice Cedex 3
France
pecman@unice.fr

SISTEMATIZACIJA NOTACIJE I ANOTACIJE KOLOKACIJA

U radu se nastoji sistematizirati notaciju i anotaciju kolokacija. Iako postoji velik broj radova posvećenih leksikologiji, leksikografiji i frazeologiji, oni su rijetko posvećeni pitanju markera koji omogućuju izdvajanje jedinica iz njihovih frazalnih konteksta ili pitanju oznaka koje određuju upotrebu tih jedinica u jeziku. Autori rječnika obično nude svoje vlastite markere i oznake kako bi zadovoljili potrebe izdavača, no njihova se upotreba rijetko sistematizira. U ovom se radu kritički ispituje upotreba markera i oznaka u leksikografiji i nude se rješenja do kojih se došlo u okviru znanstveno-istraživačkog projekta na Sveučilištu u Nici (Pecman 2004a). U zaključku se ističe važnost stvaranja sistematiziranih i racionalnih procedura za procesiranje kolokacija.