How the Notion of "Knowledge Rich Context" Can Be Characterized Today

Anne Condamines*

CLLE (Cognition, Langues, Langage, Ergonomie), CNRS and U. of Toulouse 2, Toulouse, France

This paper studies how the notion of Knowledge Rich Contexts (KRCs), initially defined as passages of text allowing access to elements that define terms, has evolved. It discusses related work in Natural Language Processing (NLP), which has led to disciplinary convergence. It mainly presents the way in which KRCs can manifest themselves within a corpus, in the form of explicit or implicit markers and lexico-syntactic or morphological markers. Extra-linguistic elements are examined through the study of the domain and the textual genre of the corpus studied, and by the adaptation of KRCs to a particular demand (evolution in time, interdisciplinarity…). All types of functioning are illustrated with real data from different corpus-based studies. The study of the notion of KRC allows us to address fundamental questions in the linguistic study of specialized texts, such as the link between form and meaning and between linguistic and extra-linguistic contexts.

Keywords: extra-linguistic context, Knowledge Rich Context, knowledge pattern, linguistic pattern, textual terminology

INTRODUCTION

The notion of Knowledge Rich Context (KRC), defined in 2001 by Ingrid Meyer, is very often used in terminology and even in knowledge engineering. Initially proposed to focus on portions of texts that could help the terminographer to define terms, this notion has been developed in many directions.

Based mainly on studies we have conducted, we show how the notion of KRC has evolved and how its characterization now takes into account both the linguistic context (the lexical environment of the supposed KRCs), and the extra-linguistic context (the communicative situation in which the exchanges take place and the specific purpose of the study).

Several elements have contributed to the evolution of the concept of KRC. The proximity of the "KRC" approach to that of information extraction has led linguists to refine their descriptions to make them automatable. The links with corpus linguistics have led terminologists to take linguistic variation into account. Finally, the needs of organizations have required the consideration of KRCs other than just relationship markers.

These three elements are discussed in the article.

In Section The Concept of Knowledge Rich Context (KRC) and Its Links With Knowledge Engineering, the article presents the relationships between corpus terminology and NLP and the developments that separated them. Natural Language Processing (NLP), and particularly Knowledge Engineering, initially found common ground in the notion of KRC. Thus, in information extraction, knowledge patterns were used to extract information from texts. Such a "converging" approach is losing ground in NLP, compared to statistical methods, but it has not totally disappeared.

Focusing only on the linguistic functioning, Section Evolution of KRCs From a Linguistic Point of View presents studies that have explored how KRCs as initially defined are manifested in discourse, i.e., which lexico-grammatical structures can be associated with which relations. It shows that KRCs can be explicit or implicit. The case of a suffix playing the role of a KRC is also detailed. Other studies show how the extra-linguistic dimension (domain and textual genre) can be taken into account to characterize the functioning of KRCs and improve the efficiency of their implementation with knowledge patterns adapted to domain and genre. Finally, the article describes how demands that go beyond the constitution of terminologies can be taken into account by adapting the notion of KRC. The KRC concept remains valid in such cases, but it must be adapted to these new needs.

THE CONCEPT OF KNOWLEDGE RICH CONTEXT (KRC) AND ITS LINKS WITH KNOWLEDGE ENGINEERING

At the beginning of the 1990s the evolution both of terminology and Natural Language Processing (NLP) allowed the emergence of KRCs (see Section Terminology and the Use of Corpora in the Early 1990s). This convergence allowed terminology to develop a more systematic and automatable approach to building terminology networks and to create Terminological Knowledge Bases (see Section Terminological Knowledge Bases).

From a methodological point of view, the comparison was especially obvious with information extraction, which was not only concerned with the search for semantic relations but also used knowledge patterns (see Section Knowledge Rich Contexts and Knowledge Patterns). NLP and terminology are now evolving separately, but meeting points are still active and productive (see Section KRCs and the Evolution of Natural Language Processing).

Terminology and the Use of Corpora in the Early 1990s

The notion of Knowledge Rich Context appeared at a particular time in the history of terminology, in connection with the evolution of corpus linguistics and the rapprochement of terminology with knowledge engineering in the early 1990s.

At this time, corpus linguistics focused mainly on the development of methods to help define words semi-automatically using local grammars (regular patterns dedicated to identifying defining contexts) (Sinclair, 1991). Among the defining elements, markers of semantic relations such as hyperonymy and meronymy were particularly explored. It seemed possible to adapt these methods that were initially defined for general language to specialized languages since markers of semantic relations such as hyperonymy or meronymy described from general corpora are also present in specialized corpora. Some researchers even specialized in definitions in scientific contexts and the linguistic devices for indicating them (Flowerdew, 1992). The term of corpus-based terminology appeared (Gamper and Stock, 1998, p. 149) and it

became possible to consider identifying relationships between terms (semi)-automatically.

For knowledge engineering, the need for structured and tool usable knowledge was crucial. Moreover, the mode of knowledge representation in the form of concept networks is particularly well-adapted to tools. As in terminology, hyperonymy and meronymy relations, as well as causal or temporal relations, are particularly present in knowledge bases representing a domain. It became clear that the possibility of using corpora as knowledge reservoirs made it easy to collect knowledge using regular linguistic patterns, at least in a first step. Collection was also more reliable, as experts do not always have the time or the skills to report their knowledge in relational relations. Hence, knowledge engineering and corpus terminology shared the same perspective. Skuce and Meyer even spoke of a "symbiotic relationship" between terminology and knowledge engineering (Skuce and Meyer, 1991).

With the development of NLP and its rapprochement with Artificial Intelligence (AI) and deep-learning, it is obvious that terminology and NLP have drifted apart. However, the impact of the latter on the former had a great importance on the latter one by allowing a more systematic description of the KRCs.

Terminological Knowledge Bases

The most tangible evidence of this rapprochement between the two disciplines was the creation of the concept of Terminological Knowledge Base (TKB) (Condamines, 2018). Compared to existing terminology databases, the major evolution of TKBs was to replace (or at least accompany) the definitions of terms with explicit relationships as in example (1):

1) *The payload is the part of the malware that performs a malicious action.*

Payload: is-part-of malware.
Has-for-function: performing a malicious action.

This concept of TKB was worked on by different multidisciplinary teams at about the same time but was first proposed by Meyer's team (Meyer et al., 1992). In the specific field of corpus-based terminology, Meyer named Knowledge Rich Contexts (KRC) the linguistic contexts that can be used to identify relations between concepts:

> "We define knowledge-rich contexts as naturally occurring utterances that explicitly describe attributes of domain-specific concepts or semantic relations holding between them at a certain point in time, in a manner that is likely to help the reader of the context to understand the concept in question" (Meyer, 2001, p. 281).

In brief, KRCs are contexts that can be used to construct relations between terms. As we have seen, the usable linguistic forms of these KRCs are structures that can be realized in different discursive forms. For example, another possible form of example (1) is

1') *The part of the malware that performs a malicious action is named payload.*

Taking into account these possible patterns linked to a relation from a corpus study is not always easy. In particular, it can be difficult to assess whether the addition of elements has an impact on the identification of a relation, for example, the addition of a quantifying adverb such as *often*, *a lot* or *rarely*, but the existence of linguistic patterns that can be associated with a given relation is part of a speaker's meta-linguistic competence.

These linguistic elements realizing KRCs have been given different names, for example "defining expositive" (Pearson, 1998) or "linguistic signalling devices" *(*Flowerdew, 1992), or even "diagnostic frame" (Cruse, 1986). Note that these naming choices imply two different types of functioning. The first two (*defining expositive* and *linguistic signaling devices*) are rather oriented toward the speaker, who wishes to guide the linguistic interpretation of his/her message. The third (*diagnostic frame*) is rather oriented toward an analyst/interlocutor wishing to "diagnose" possible semantic interpretations. We will come back to these two aspects in Section Speakers vs. Analysts: the Case of Epilinguistic Patterns.

Knowledge Rich Contexts and Knowledge Patterns

As already mentioned, in order to be used, knowledge rich contexts have to be expressed in the form of local grammars (Barnbrook and Sinclair, 2001; León-Araúz, 2014). Such grammars are close to the knowledge patterns used in NLP. Obviously, this aspect helped to bring terminology and NLP closer together, at least initially.

Information Extraction

The purpose of using KRCs is very similar to that of information extraction, even if the latter is not concerned only with the extraction of semantic relations.

> "Information extraction is a subfield of natural language processing that is concerned with identifying predefined types of information from text. [...]. Most information extraction systems use some form of extraction pattern to identify potentially relevant information" (Riloff, 1999, p. 436).

In the case of terminology, the relations are rather stable knowledge in the domain and rather ontological (describing the objects of the domain). In the case of information extraction, each new search requires identifying the regular linguistic means of accessing the information sought (for example, a search for the size of the highest towers in the world). However, extraction patterns and local grammars have the following aims in common: (a) to predefine the types of information to be searched for (conceptual relations in the case of KRCs), and (b) to define a priori the linguistic patterns that can be used to search for this information. Like extraction patterns, local grammars should be sytematizable and automatable. The first knowledge patterns, for classical relations such as hyperonymy (generic-specific) or meronymy (part-of), were defined by introspection [for example, for hyperonymy: [Y is the most adj of X], [X excerpt Y], [Y be an X that…]]:

(2) *The blue whale is the most massive of all mammals.*

In (2), we understand that [blue whale] has [mammal] as its generic (hyperonym): [blue whale] is-a [mammal].

Knowledge patterns, as parts of the metalinguistic functioning of the language, are present in all languages. Numerous studies have been conducted in different languages such as English, French, Spanish, Catalan, Russian… To cite but a few: (Nuopponen, 1994; Cabré et al., 2001; Feliu and Cabré, 2002; Barrière, 2004; L'Homme and Marshman, 2006; Auger and Barrière, 2008; Schumann, 2011; León-Araúz et al., 2016). The most frequently studied relationships are hyperonymy, meronymy and cause.

Though Meyer was aware of greater possibilities, the notion of Knowledge Rich Context as she initially defined it has three main characteristics:

i. In *Knowledge rich context*, *context* concerns only the linguistic/textual environment of a term or of two related terms. Note that, conversely, in psychology for example, this same expression is used most of the time to refer strictly to the extra-linguistic context [see for example the use of this term in a paper in developmental psychology entitled: *The development of scientific reasoning in knowledge-rich contexts* (Schauble, 2016)].

ii. Only the identification of relations or characteristics about concepts is concerned by the notion of knowledge rich context.

iii. Only lexico-syntactic elements are considered as KRCs, i.e., (portions of) sentences.

These three aspects of the definition of KRCs have been worked on and developed from different perspectives. These new perspectives are presented in part 3.

KRCs and the Evolution of Natural Language Processing

Since the early 1990s, NLP researchers have tried to assist the search for relationship patterns and, in particular, to make it no longer dependent on introspection alone. Hearst's study (Hearst, 1992) was a driving force in this respect. The originality of Hearst's approach is that it combines top-down and bottom-up approaches using a corpus of the domain. First, knowledge patterns are used to identify terms linked by the hyperonymy relation. Then, when pairs of terms have been thus identified, they are reprojected onto the corpus in order to make new hyperonymy patterns emerge from the corpus. There is a permanent back and forth between a priori knowledge and the knowledge present in the corpus. NLP studies have aimed to automate this double movement, as presented in Morin and Jacquemin (2004) for example.

The so-called endogenous approach, which aims to bring out linguistic regularities from the corpora themselves, has developed considerably in recent years, thanks to the availability of increasingly large volumes of data and the learning methods developed by Artificial Intelligence (Buitelaar and Cimiano, 2008). Most often, it is not mainly a matter of discovering what relationship exists between two terms but rather that the same relationship exists between several pairs of terms because

they appear in similar contexts. These approaches are based on the proposals of the distributional linguists of the 1950s often summed up in Firth's famous dictum: "you shall know a word by the company it keeps" (Firth, 1957, p. 11). For a number of applications (e.g., information retrieval, which is about finding the document that relates to the user's topic of interest), it is not necessary to know what the relationship is between two terms, but it is very important to know that these terms are related by the same relation as other pairs of terms.

Nowadays, most of these tools do not use *a priori* linguistic knowledge but rely only on statistical methods called neural approaches. More rarely, some studies combine statistical and symbolic approaches (Rojas-Garcia and Cabezas-Garcia, 2017) or even compare the two approaches (Roller et al., 2018).

Due to this evolution of NLP research, the terminology and knowledge engineering communities have moved apart to some extent (Condamines and Picton, (2022)). Howver, the purpose of this article is not to go into this aspect but rather to focus on the linguistic evolution.

EVOLUTION OF KRCS FROM A LINGUISTIC POINT OF VIEW

The possibilities of automatic processing but also the evolution of demands concerning the management of specialized documentation had a major role on the diversification of KRC studies. In this section, we show how, even with these new perspectives, the notion of KRC is crucial because KRCs play a pivotal role both in accounting for discourse and lexical variations of the same pattern, and in taking into account the extra-linguistic context (communicative situation and specific needs related to specialized documents). Part Unpredictable Linguistic Achievements of KRC focuses on unpredictable KRCs. Part KRCs and Translation discusses how translation (as a process or as an outcome) informs the notion of KRC. Section Taking Into Account the Extra-Linguistic Context presents examples of KRC variation in relation to external demands and the impact on the notion of knowledge wealth.

Unpredictable Linguistic Achievements of KRC

The first KRCs were identified by introspection. With the use of AI learning methods, new markers emerged that were much less predictable but corresponded to identifiable usages in general and large corpora. However, some language elements can be used much more rarely to identify semantic relations, without being directly associated with these relations (Ssee Section Speakers vs. Analysts: The Case of Epilinguistic Patterns). Other elements below word level, such as affixes, can also be used as markers of relations (see Section Other Linguistic Elements in the Role of KRCs. What About Affixes? The Case of Denominal Adjectives).

Speakers vs. Analysts: The Case of Epilinguistic Patterns

As we have seen, the main idea associated with the notion of KRC is related to the fact that some language elements can be

regularly associated with a stable interpretation. Nevertheless, if we take the case of a linguist/terminologist who looks for relation patterns in corpora, the situation is rather different.

Indeed, some elements that *a priori* are not intended to be knowledge patterns can become so. They do not really function as linguistic signs but rather as cues specific to the textual domain or genre. Following Culioli's proposals, "classical" KRCs can be described as metalinguistic and the patterns discussed in this section as epilinguistic (Condamines, 2017):

> "The difference between metalinguistic and epilinguistic activity is that the former is deliberate whereas the latter is an uncontrolled inner process" (Culioli, 1979, p. 205).

We encountered this case in a very specific field and genre: didactic texts in the field of natural sciences, with the preposition *chez (in)* (Condamines, 2000). Thanks to a comparative study of several corpora, we showed that in corpora belonging to this genre and domain, *chez* could be used to identify a meronymic relation in almost 50% of cases, as in (3) and (4).

3) Chez la plupart des Lorisiformes, il existe des zones glandulaires circumgénitales.
   In most Lorisiformes, there are circumgenital glandular areas.
4) *Chez les colobinés, le nez fait saillie sur la lèvre supérieure.*
   In the Colobinés *the nose juts out over the upper lip.*

Of course, *chez* by itself does not express meronymy. The preposition comes from the Latin word *casa* (in the home of). *Chez* is always followed by an animate; it is impossible for example to find occurrences such as *chez le carbone* (in carbon). What is specific in the case of didactic genres in natural sciences is that the phrase containing *chez* focuses on an animal or a plant and the rest of the sentence gives information about the animal or the plant concerned, such as eating habits, habitat, reproductive mode, etc. as in (5):

5) *Chez le fraisier, la préfloraison est calvaire pour le calice, quinconciale pour la corolle.*
   In the strawberry plant, pre-flowering is in calvary for the cup, and in staggered rows for the corolla

But in a very large number of cases, this information concerns parts of the animal/plant.

Out of context, it would be impossible to predict that *chez* is a marker (or even a cue) of meronymy, but in the precise case of didactic genres in natural science, if one wants to identify relations between terms, this preposition may be used as a knowledge cue, in particular to access meronymic knowledge.

Other Linguistic Elements in the Role of KRCs. What About Affixes? The Case of Denominal Adjectives

Terminologists are seldom interested in the morphological level, below the word. We can however note some studies that have been carried out, especially in an NLP perspective (Namer and Zweigenbaum, 2004 for example).

Among morphologists, there is a lively debate about whether affixes (prefixes and suffixes) are really linguistic signs or whether

they are only a means to change the grammatical category from noun to adjective (Rainer, 2013). *Are affixes signs?* Lehrer asks in the title of her article (Lehrer, 2000). In agreement with the proponents of constructional morphology, she argues that they are:

Morphological and syntactic structures can be quite similar, and accounted for by constructional schema (Booij, 2016).

We suggest that some affixes can be used as relation markers because the underlying polysemy is lifted in some fields, as is often the case for terms.

We will focus here on denominal adjectives, sometimes called relational adjectives, because they may be of particular interest for terminologists. In French, in [Nom adjectif relationnel (Relational Adjective Noun)] structures, we can identify two nouns: the head noun and the base noun used to constitute the adjective with the addition of a suffix. For example in *ville côtière (coastal town),* there are two nouns: *côte (coast)* and *ville (town)* and a relational suffix *-ier* (-al) indicating a relation between the two nouns.

Out of context, most suffixes are polysemous. Furthermore, as with lexico-syntactic patterns, domain often plays a role in the interpretation of suffixes. We conducted a study on the adjective *volcanique* (volcanic) in a French corpus on volcanology (Condamines, 2020). The adjective derives from the noun *volcan*. The corpus contains 730 000 words and [N *volcanique*] occurs more than 2 020 times.

In the online French dictionary *Le Trésor de la Langue Française*, three meanings are given for *volcanique*:

Sens 1: Qui constitue le volcan, qui est de la nature du volcan.

Meaning 1: Which constitutes the volcano, which is of the nature of the volcano.

Sens 2: Qui provient d'un volcan, est expulsé par un volcan, a pour origine un volcan.

Meaning 2: Which comes from a volcano, is expelled by a volcano, has for origin a volcano.

Sens 3 (figuré): Qui présente les caractères d'un volcan; impétueux, dont les réactions sont imprévisibles.

Meaning 3 (figurative): Which has the characteristics of a volcano; impetuous, whose reactions are unpredictable.

Of course, the meaning that interests us is the first one. But the second one may also be interesting as the elements that are ejected from the volcano often fall back onto the volcano and become external parts of it.

What the corpus analysis shows is that in structures containing *volcanique*, there may be a meronymic relationship between the head noun and *volcan* [as in *event volcanique* (*volcanic vent*)] or between *volcan* and the head noun [as in *massif volcanique* (*volcanic massif*)]. In some cases, such as *activité volcanique* (*volcanic activity*), the relation is not a meronymy. In almost 60% of cases, however, the relation is meronymic.

To compare structures with relational adjectives and structures with lexico-syntactic markers we also looked for lexico-syntactic patterns expressing a meronymic relationship between the head noun and *volcan*. We found well-identified

TABLE 1 | Morphological and lexico-syntactical patterns for *volcan*.

N *volcanique(s)*	215
N *volcanique(s)* with a meronymy	126
N lexico-syntactic *volcan(s)* with a meronymy	124
N present in the two structures with a meronymy	87

patterns such as [(être) constitué de], [comporter], [(être) composé de]

6) *Il s'agit d'un volcan, culminant à 1 397 m,* constitué d'un *empilement de laves et de tephra*
   It is a volcano, culminating at 1 397 m, consisting of *a pile of lava and tephra*

7) *Ce volcan est actif depuis le Pléistocène supérieur et* est composé d'*une caldeira et d'une quarantaine de cônes*
   This volcano has been active since the Upper Pleistocene and is composed of *a caldera and about forty cones*

We also identified patterns specific to the corpus such as [(être) parsemé/hérissé de ((to be) dotted/bristling with)]

8) *L'île du Nord de Nouvelle-Zélande* est parsemée de *volcans*
   The North Island of New Zealand is dotted *with volcanoes*

9) *L'Enclos est le fond d'un volcan ancien* hérissé de *scories,*
   L'Enclos is the bottom of an ancient volcano bristling with *slag*

In the corpus, *volcanique(s)* (*volcanic*) is used with 215 different nouns. Of these 215 nouns, 126 show a clear meronymy (*cheminée volcanique, aire volcanique*) and 89 show no meronymy (*activité volcanique, épisode volcanique*).

Concerning lexico-syntaxic patterns, 124 nouns related to *volcan* were found.

Eighty seven nouns appear with *volcan* (*volcano*) both in a structure with a meronymic knowledge pattern and a structure containing *volcanique* (N *volcanique*).

Table 1 presents the results obtained.

If the aim is to identify relations between terms, one might think that morphological markers and lexico-syntactic markers are redundant. From a semantic point of view, this is not exactly the case. In fact, the lexico-syntactic markers are chosen to give explicit information, whereas the morphological markers refer to supposedly known information. However, once again, from the point of view of the linguist-terminologist, it appears that using the suffix -ique as a relation marker, at least for *volcanique*, can be effective.

The meronymic interpretation is less reliable with suffix markers than with lexico-syntactic markers (even if it may depend on the lexical marker concerned). However, one of the advantages of suffixes is that they vary much less than lexico-syntactic markers (which can be accompanied by adverbs, be dislocated etc.). Morphological markers are therefore easier to identify. In the case of *volcan,* it would probably be best to start by identifying the N *volcanique*(s) structures and then use the structures with lexico-syntactic markers. However, further studies would be needed to confirm this approach, using other nouns with—*ique* in the same domain (where they are

very numerous, for example *andésitique (andesitic), cratérique (crateric), pyroclastique (pyroclastic)* etc.) and in other domains. The study is in progress.

It should be noted that with the use of morphemes as relation markers, the notion of KRC evolves since the marker is inside a word and this word is often part of a structure which is itself a term. It is known that in about two thirds of cases terms are made up of several words. We end up with a structure where terms and KRC are closely intertwined, for example in *roche volcanique*, there are three terms: *roche (rock), volcanique (volcanic)* and *roche volcanique (volcanic rock)* and a KRC: *-ique*. Moreover, *roche volcanique (volcanic rock)* can, in turn, be integrated into a phrase with a KRC such as: *les minéraux constitutifs des roches volcaniques (the constituent minerals of volcanic rocks)*. And it is not inconceivable that *minéraux constitutifs des roches volcaniques* is itself a term!

Few linguistic studies have been conducted to systematically describe the use of affixes as KRC. This is a promising avenue, however, which again positions KRCs as pivotal elements, both within texts themselves since they link terms, and between the linguistic and the extra-linguistic context since they too vary according to textual domain and genre.

KRCs and Translation

It is particularly interesting to consider translation in order to describe the functioning of KRCs. From the perspective of translation assistance, studies have shown that the most widely used KRCs are not necessarily conceptual relation markers (see Section KRCs in a Translation Assistance Task). From the point of view of the translation result, it can be seen that translations of KRCs can be used to provide new elements for building a terminology network (see Section Contribution of the Translation of KRCs to Building Terminological Networks).

KRCs in a Translation Assistance Task

Some studies have focused on the role of KRCs in a translation task (Bowker, 2012; (Condamines et al., 2013). In Picton and Josselin-Leray (2019), the authors considered two kinds of KRCs: Conceptual KRCs and Linguistic KRCs.

Terminology and translation have long followed a common path. And often, terminology databases have been built with the idea of a translator user in mind. While this was not the idea behind Meyer's proposal with KRCs, it was legitimate for the notion of richness suggested by KRCs to be evaluated for the task of translation itself. As part of the research project CRISTAL (Contextes Riches en Connaissance pour la Traduction), an experiment was conducted in order to examine how translators use (or could use) KRCs. Picton and Josselin-Leray (2019) presented the results of experiments conducted with 42 participants to find out which KRCs are really used by translators, when they use them, and for which terms. They used a mixed-methods approach comprising both quantitative and qualitative aspects. They examined 5 types of data: logs that take into account the participants' actions, a questionnaire after the translation task, interviews with users, videos, and transcripts of Cue-based retrospective Verbalization tasks. They presented two types of KRCs that could be useful for translators: linguistic

KRCs (concerning collocations) and conceptual KRCs (the ones considered by Meyer, 2001). The results show that the notion of KRCs for translators is more complex to define than that of KRCs for terminologists. The richness of KRCs varies according to the moment in the translation process and does not only concern conceptual knowledge but also, and sometimes in the same context, linguistic richness (collocation), hence the notion of linguistic KRCs that the authors proposed.

Contribution of the Translation of KRCs to Building Terminological Networks

It is also interesting to study how the translation of KRCs can be used in order to improve the constitution of a terminology network.

In the field of translation, Baker (1996) identified 4 strategies used by translators:

- "explicitation," i.e., "a tendency to spell things out rather than leave them implicit in translation"
- "simplification," i.e., "the tendency to simplify the language used in translation"
- "normalisation," i.e., "the tendency to exaggerate features of the target language and to conform to its typical patterns"
- "levelling out" i.e., "the tendency of translated text to gravitate toward the center of a continuum."

The phenomenon of explicitation can be particularly relevant for a terminologist working with an aligned corpus, with the objective of building a terminology network. This is what we found when we studied the hyperonymy patterns in a corpus translated from French into Italian (Condamines et al., 2021). We used an aligned corpus composed of articles from the French monthly newspaper, *Le Monde Diplomatique*, translated into Italian. The 1-million-word corpus concerns the field of political economy. This monthly can be considered as semi-popularization (from experts to semi-experts). As mentioned above, this textual genre is known to be rich in knowledge markers. In her master thesis, Federzoni (Federzoni, 2018) projected 39 hyperonymic patterns in both languages to see what contexts corresponded to them in the other language. French patterns were listed in the PhD thesis by Lefeuvre (2017) and constitued the MAR-REL resource (http://redac.univ-tlse2.fr/misc/mar-rel_fr.html). Italian patterns were listed by translating French patterns. Without going into detail, we can say that in 16% of cases, a hyperonym was added in the translation, for example:

10) *[…] Classés au centre gauche, La Repubblica et le Corriere della Sera […] I due più grandi giornali della penisola, la Repubblica e il Corriere della Sera.*

   [*[…] Classified as center-left, La Repubblica and Corriere della Sera,* the two largest newspapers in the peninsula […]]

This phenomenon may be explained by the fact that the translator not only translates literally but also provides additional information to help the reader unfamiliar with the culture of the country.

The addition of hyperonyms is particularly interesting when building a terminology network as it can be used to attach bits of

networks. It is assumed that the conceptual network underlying a text and its translation is the same. If this network emerges in different forms in different languages, this variety enriches the possibilities of accessing the network. In the same way, translation can provide access to new patterns that had not been identified in one or the other language.

Taking Into Account the Extra-Linguistic Context

In the early 1990s, it was uncommon to consider variation in terminology and, in particular, variation linked to the extra-linguistic situation. What was searched for was rather stability, as advocated by Wüster and his General Theory of Terminology (GTT) in the 1930s (Wüster, 1979). In this theory, it was more a question of prescribing a reference terminology in order to make exchanges more transparent. Nevertheless, descriptive approaches developed since 1990 based on the analysis of real uses have played an important role in terminology studies, for example, the Communicative Theory of Terminology (Cabré, 2009), Socioterminology (Gaudin, 1993), Sociocognitive Terminology (Temmerman, 2000), Textual Terminology (Condamines and Picton, (2022); Faber and L'Homme, (2022)). The confrontation with variations and also new needs that rely on terminology became inevitable (Condamines, 1995). Section KRCs, Domain, and Genre presents the issue of taking into account the context of communication via the notions of domain and genre. Section Wealth of Knowledge shows how extra linguistic needs influenced the definition of KRCs.

KRCs, Domain, and Genre

Various corpus studies have shown variation in the presence and/or interpretation of KRCs depending on the domain and textual genre of the texts studied. In Marshman et al. (2008), for instance, the authors studied the "portability" of KRCs between domains and other authors have explored the role of the textual genre in the presence and interpretation of KRCs (Condamines, 2002, 2008). In such studies, it is the extra-linguistic dimension that is taken into account, in particular the presence of a "defining intention" in certain communication situations producing texts belonging to certain genres. For example, when experts talk to novices or semi-experts, they use knowledge patterns to make their point more explicit and their discourse contains more KRCs (Pearson, 1998). Conversely, scientific communication between peers is KRC-poor because one of the characteristics of specialists is that, belonging to the same community, they share knowledge that does not need to be discursively defined. As for the domain, it has an impact on the presence of certain relationships. For example, the medical domain leads to a strong presence of is-a-symptom-of and hence patterns linked to this relation. Domain and/or genre also limit the polysemy of some patterns. For example, in a previous study we showed that *avec, with* or *con* (respectively, in French, English and Spanish) in nominal groups express a meronymic relation in almost all occurrences in real estate classified ads (Condamines, 2009). These prepositions are then used to emphasize the whole of which the part is specified. For example in:

11) *living room with large veranda,*

it is clear that the veranda (a large one, moreover) gives value to the living room.

Such examples show that it is difficult to completely isolate the interpretation of texts, especially specialized texts, from their context of writing and even from their context of interpreting (see Section Wealth of Knowledge). In fact, the interpretation of texts is not unique but may depend on the purpose of their analysis.

Wealth of Knowledge

Corpus terminology was initially motivated by the possibility of constructing terminology networks and, consequently, of systematizing the definition of terms. But the increase in the amount of textual data made available has been accompanied by new needs, especially in public or private organizations. This evolution has had an impact on the notion of "wealth of knowledge". For a linguistic context, being rich in knowledge depends on both the supposedly shared linguistic knowledge of the speaker and the interlocutor (using the same metalinguistic knowledge), and on their intentions, which are not necessarily identical. In addition, linguists/terminologists are not just readers of the texts, they use corpora that were not written for them. When they look for knowledge rich contexts to build networks of terms, they focus on portions of text that are usable for the construction of relations. But they may also be interested in other types of knowledge rich contexts, depending on their objective.

Examples of Various KRCs Depending on the Objective of the Study

As we have seen, organizations sometimes have other needs than just building a reference terminology. This was the case with three requests from the CNES: Center National d'Etudes Spatiales (French National Center for Spatial Studies). In each study, we used the same approach, based on textual terminology (Condamines and Picton, (2022)), i.e., comparing the lexical functioning of corpora constituted according to a study objective. The results of these studies are briefly presented in the following 3 sections.

Taking Into Account the Evolution Over Time

In the case of "Evolution over time," the question was to identify changes or evolution in the meaning of the terms in corpora from different periods. If the evolution is identified by identifying an evolution of the contexts of occurrence of the terms (their distribution), then in certain cases, KRCs can be used. These are cases where the speakers are aware of this evolution and mark it in the text (Picton, 2014). This emergence is marked for example by the adjective *nouveau/new* or a verb such as *apparaître/appear* in (12):

12) *Un nouveau produit est apparu sur le marché il y a quelques années, c'est le multi-barettes. A new product appeared on the market a few years ago, namely the multi-bar*

The Role of Interdisciplinarity on the Use of Terms

The issue of interdisciplinarity concerned the study of meaning shifts between disciplines. At the request of CNES, we worked on a new discipline, exobiology, which studies life outside the solar system. It involves astronomy, chemistry, biology and geology. The aim was to see how each discipline borrows terms from the others and how these borrowings, often associated with meaning shifts, contribute to creating new concepts (Condamines, 2014).

As for the evolution over time, this study was mainly based on the study of the modifications in the contexts of the terms depending on the discipline. But it was also possible to use KRCs related to the issue, i.e., contexts where speakers were aware of a shift in meaning in their discipline and expressed it through explicit formulations. Thus, in (14):

13) *Du point de vue chimique, une oxydation est définie comme la perte d'un ou plusieurs électrons par un atome ou une molécule*

　　From a chemical point of view, an oxidation is defined as the loss of one or more electrons by an atom or a molecule

This excerpt signifies that the proposed definition is appropriate for chemistry but could be different in another field.

Characterization of the Despecialization of Terms in the General Language

In the case of despecialization, CNES wanted to know how the space domain permeates the knowledge of the general public. We approached this question by studying determinologization, i.e., the fact that a term enters the vocabulary shared by the speakers of a language. As in previous studies, KRCs were used to identify this phenomenon. Thus, in (14):

14) *phénomènes aérospatiaux non identifiés (PAN), plus connus du grand public sous le terme d'ovni*

　　unidentified aerospace phenomena (UAP), better known to the general public under the term UFO

the term *grand public/general public* and more specifically *better known to the general public under the term* makes it possible to establish a link between the scientific term and the popularized term. As such, they play the role of KRCs.

The common characteristics of all these KRCs are that they are linked to a predefined objective and that they can be identified by introspection, as for the initial KRCs. Moreover, in most cases, there is a trigger element (a clue) which, completed by another element of the context, allows the interpretation. Taking into account the extra-linguistic context allows us to refine our understanding of the functioning of KRCs and to adapt the effectiveness of studies to increasingly varied needs.

CONCLUSION

By proposing to consider some portions of texts as Knowledge Rich Contexts, Meyer introduced the idea that some language elements could be given a salient role in order to perform a particular task in specialized domains, namely building a terminological network.

But salience may concern either linguistic phenomena themselves or the perspective of a reader who has a particular goal in mind. As noted by Boswijk and Coler,

> "[…] in cognitive linguistics, something is salient either because of an intrinsic property (bottom-up salience) or because of an external contextual factor (top-down salience)" (Boswijk and Coler, 2020, p. 717).

The KRCs defined by Meyer correspond to bottom-up salience. They can be enriched with elements below the word (affixes) or with non-metalinguistic elements playing the role of cues in certain textual domains and genres (epilinguistic markers), making it difficult to draw up an exhaustive list of KRCs.

Extra-linguistic elements can intervene in two ways. They can play a role in the variations of linguistic patterns, in relation to the domain concerned or the textual genre. Extra-linguistics can also intervene in the form of particular needs that lead to the definition of KRCs designed to meet these needs. Whatever the nature of the KRCs, linguistic elements that may correspond to them can be identified by introspection in a first step. KRCs that are specific to the corpus studied and therefore often unpredictable can then be identified in a second step.

KRCs appear at a pivotal point between linguistic and domain-related knowledge and between linguistic analysis and extra-linguistic needs. To take into account the extra-linguistic context, the constitution of the corpus is crucial since it must take into account, on the one hand, the elements of characterization of domain and genre and, on the other hand, the objective of the study. While terms are a determining entry point into specialized texts, it is the KRCs that must be mobilized in order to meet analytical needs. In the future, we can think that a reflection on the efficiency of KRCs (in connection with their implementation) will require taking into account different parameters: the role of the domain and the textual genre, the position of the KRCs, i.e., intra or extra term (morphology vs. syntax), the need justifying the use of KRCs (to check coherence, to control the evolution of knowledge…), the task to be accomplished (to constitute a terminological network, help translators…). These descriptions will undoubtedly be assisted by AI methods, but fundamentally, in order to understand the links and complementarities between all these parameters, linguists certainly still have a long way to go.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Auger, A., and Barrière, C. (2008). Pattern-based approaches to semantic relation extraction: a state-of-the-art. *Terminology* 14, 1–19. doi: 10.1075/term.14.1.02aug

Baker, M. (1996). "Corpus-based translational studies: the challenges that lie ahead," in *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, ed. H. Somers (Amsterdam; Philadephia: John Benjamins), 175–186. doi: 10.1075/btl.18.17bak

Barnbrook, G., and Sinclair, J. (2001). "Specialised corpus, local and functionnal grammars," in *Studies in Corpus Linguistics*, eds. M. Ghadessy, A. Henry, and R. L. Roseberry (Amsterdam; Philadephia: John Benjamins Publishing Company), 237–75. doi: 10.1075/scl.5.15bar

Barrière, C. (2004). Knowledge-rich contexts discovery. *Lect. Notes Comput. Sci.* 3060, 187–201. doi: 10.1007/978-3-540-24840-8_14

Booij, G. (2016). "Construction morphology,"in *The Cambridge Handbook of Morphology*, eds A. Hippisley and G. Stump (Cambridge: Cambridge University Press), 424–448. doi: 10.1017/9781139814720.016

Boswijk, V., and Coler, M. (2020). What is salience? *Open Linguist.* 6, 713–722. doi: 10.1515/opli-2020-0042

Bowker, L. (2012). "Meeting the needs of translators in the age of e-lexicography: exploring the possibilities," in *Electronic Lexicography*, eds S. Granger and M. Paquot (Oxford: Oxford University Press), 379–387. doi: 10.1093/acprof:oso/9780199654864.003.0018

Buitelaar, P., and Cimiano, P. (2008). *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*. Amsterdam: IOS Press.

Cabré, M.-T. (2009). The Communicative Theory of Terminology, a linguistic approach of terms. *Rev. Française Lingusit. Appl.* 2, 9–15 doi: 10.3917/rfla.142.0009

Cabré, M.-T., Morel, C., and Tebé, C. (2001). "Propuesta metodológica sobre cómo detectar las relaciones conceptuales en los textos a través de una experimentación sobre la relación causa-efecto," *La terminología científico-técnica: Reconocimiento, análisis y extracción de información formal y semántica*, eds M. Teresa Cabré and J. Feliu (Barcelona: Institut universitari de lingüística aplicada, Universitat Pompeu Fabra), 165–170.

Condamines, A. (1995). Terminology: new needs, new perspectives. *Terminology* 2, 219–238. doi: 10.1075/term.2.2.03con

Condamines, A. (2000). *Chez* dans un corpus de sciences naturelles: un marqueur de méronymie ? *Cahiers Lexicol.* 77, 165–187. doi: 10.15122/isbn.978-2-8124-4329-9.p.0167

Condamines, A. (2002). Corpus analysis and conceptual relation patterns. *Terminology* 8, 141–162. doi: 10.1075/term.8.1.07con

Condamines, A. (2008). Taking *Genre* into account for analyzing conceptual relation patterns. *Corpora* 8, 115–140. doi: 10.3366/E1749503208000129

Condamines, A. (2009). L'expression de la méronymie dans les petites annonces immobilières; comparaison français, anglais, espagnol. *J. French Lang. Stud.* 19, 3–23. doi: 10.1017/S0959269508003554

Condamines, A. (2014). How can linguistics help to structure a multidisciplinary neo-domain such as exobiology? *BIO Web Conf.* 2, 06001. doi: 10.1051/bioconf/20140206001

Condamines, A. (2017). "Identification de relations conceptuelles en corpus spécialisés: mise en saillance définitoire métalinguistique vs épilinguistique," in *Langue française mise en relief(s). Aspects grammaticaux et discursifs*, eds M. Bilger, L. Buscail and F. Mignon (Perpignan: Presses Universitaires de Perpignan), 199–211.

Condamines, A. (2018). "Terminological knowledge bases," in *The Routledge Handbook of Lexicography*, ed. P. Fuentes (London: Routledge), 335–349. doi: 10.4324/9781315104942-22

Condamines, A. (2020). Morphologie vs syntaxe dans le repérage de relations conceptuelles en corpus spécialisés: étude comparée de N *volcanique(s)* vs N [structure lexico-syntaxique] *volcan* dans un corpus de volcanologie. *Congrès Mondial Linguist. Française.* doi: 10.1051/shsconf/20207805001. Available online at: https://www.shs-conferences.org/articles/shsconf/abs/2020/06/shsconf_cmlf2020_05001/shsconf_cmlf2020_05001.html

Condamines, A., Escoubas, M.-P., and Federzoni, S. (2021). "Apport de la traduction dans l'analyse des marqueurs de relations conceptuelles. Une étude en corpus aligné français-italien," in *Des corpus numériques à l'analyse*

linguistique en langues de spécialité*, eds C. Frérot and M. Pecman (Grenoble: Presses de l'UGA), 313–333. doi: 10.4000/books.ugaeditions.24280

Condamines, A., and Picton, A. (2022) Textual terminology: origins, principles and new challenges. *Theoretical Approaches to Terminology*, eds P. Faber and M.-C. L'Homme (Amsterdam; Philadelphia: John Benjamins).

Condamines,., A., Josselin-Leray, A., Fabre, C., Lefeuvre, L., Picton A, and Rebeyrolle, J. (2013). Using comparable corpora to characterize knowledge-rich contexts for various kinds of users: preliminary steps. *Proc. Soc. Behav. Sci.* 95, 581–586. doi: 10.1016/j.sbspro.2013.10.685

Cruse, D. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.

Culioli, A. (1979). *Why Teach How to Learn to Teach What is Best Learnt Untaught Cahiers Charles V Année*. Université Paris 7, 199–210. doi: 10.3406/cchav.1979.904

Faber, P., and L'Homme, M.-C., (eds.) (2022). *Theoretical Approaches to Terminology*. Amsterdam; Philadelphia: John Benjamins.

Federzoni, S. (2018). *Étude de la présence et du fonctionnement des marqueurs de relations conceptuelles dans un corpus aligné franco-italien* (Master's Dissertation). University of Toulouse Jean Jaurès, Toulouse, available online at: http://dante.univ-tlse2.fr/9090 (accessed April 26, 2022).

Feliu, J., and Cabré, M. T. (2002). "Conceptual relations in specialized texts: new typology and extraction system proposal," in *TKE'02, 6th International Conference in Terminology and Knowledge Engineering* (Nancy, France: Université de Nancy), 45–49.

Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.

Flowerdew, J. (1992). Definitions in science lectures. *Appl. Linguist.* 13, 202–222. doi: 10.1093/applin/13.2.202

Gamper, J., and Stock, O. (1998). Corpus-based terminology. *Terminology* 5, 147–160. doi: 10.1075/term.5.2.05gam

Gaudin, F. (1993). *Pour une socioterminologie - Des problèmes sémantiques aux pratiques institutionnelles*. Rouen: Publications de l'Université de Rouen.

Hearst, M. A. (1992). "Automatic acquisition of hyponyms from large text corpora," in *Proceedings of the 14th International Conference on Computational Linguistics (COLING'92)* (Nantes, France: Université de Nantes), 539–545,. doi: 10.3115/992133.992154

Lefeuvre, L. (2017). *Analyse des Marqueurs de Relations Conceptuelles en Corpus Spécialisé: Recensement, Évaluation et Caractérisation en Fonction du Domaine et du GenreTextuel* (Unpublished PhD dissertation). Toulouse: Université of Toulouse-JeanJaurès.

Lehrer, A. (2000). "Are affixes signs?: the semantic relationships of English derivational affixes," in *Morphological Analysis in Comparison. Current Issues in Linguistic Theory*, eds W. U. Dressler, O. E. Pfeiffer, M. A. Pöchtrager, and J. R. Rennison (Amsterdam; Philadelphia: John Benjamins), 143–154. doi: 10.1075/cilt.201.08leh

León-Araúz, P. (2014). "Semantic relations and local grammars for the environment," in *Formalising Natural Languages with NooJ 2013*, eds. S. Koeva, S. Mesfar, and M. Silberztein (Newcastle-upon-Tyne: Cambridge Scholars Publishing), 87–102.

León-Araúz, P., San Martín, A., and Faber, P. (2016). "Pattern-based word sketches for the extraction of semantic relations," in *Proceedings of the 5th International Workshop on Computational Terminology* (Osaka, Japan), 73–82.

L'Homme, M.-C., and Marshman, E. (2006). "Terminological relationships and corpus-based methods for discovering them: an assessment for terminographers," in *Lexicography, Terminology, and Translation. Text-based studies in honour of Ingrid Meyer*, ed. L. Bowker (Ottawa: University of Ottawa Press), 67–80. doi: 10.2307/j.ctt1ckpgs3.8

Marshman, E., L'Homme, M.-C., and Surtees, V. (2008). Portability of cause-effect relation markers across specialised domains and text genres: a comparative evaluation. *Corpora* 3, 141–172. doi: 10.3366/E1749503208000130

Meyer, I. (2001). "Extracting knowledge-rich contexts for terminography: a conceptual and methodological framework," in *Recent Advances in Computational Terminology*, eds. D. Bourigault, M.-Cl. L'Homme, and C. Jacquemin (Amsterdam; Philadelphia: John Benjamins), 279–302. doi: 10.1075/nlp.2.15mey

Meyer, I., Skuce, D., Bowker, L., and Eck, K. (1992). "Towards a new generation of terminological resources: an experiment in building A terminological knowledge base," in *Proceedings of the Fifteenth Conference on Computational*

Linguistics (COLING'92) Vol. 3, ed. C. Boitet (Stroudsburg, PA: Association for Computational Linguistics (ACL)), 956–930. doi: 10.3115/992383.992410

Morin, E., and Jacquemin, C. (2004). Automatic acquisition and expansion of hypernym links. *Comput. Hum.* 38, 363–396. doi: 10.1007/s10579-004-1926-2

Namer, F., and Zweigenbaum, P. (2004). "Acquiring meaning for French medical terminology: contribution of morphosemantics," in *Eleventh MEDINFO International Conference* Amsterdam: IOS Press, 535–539.

Nuopponen, A. (1994). Causal relations in terminological knowledge representation. A paper in the Symposium "Terminology and Knowledge Engineering: Knowledge representation and object description", University of Vienna, Austria, 22-23.1.1993. *Terminol. Sci. Res.* 5, 36–44.

Pearson, J. (1998). *Terms in Context*. Amsterdam; Philadelphia: John Benjamins.

Picton, A. (2014). "The dynamics of terminology in short-term diachrony: a proposal for a corpus-based methodology to observe knowledge evolution," in *The Dynamics of Culture-bound Terminology in Monolingual and Multilingual Communication, Terminology and Lexicography Research and Practice*, eds. R. Temmerman and M. Van Campenhoudt (Amsterdam; Philadephia: John Benjamins), 159–182. doi: 10.1075/tlrp.16.09pic

Picton, A., and Josselin-Leray, A. (2019). "A mixed-methods approach to characterize Knowledge-Rich Contexts for specialized translation," in *New Challenges for Research on Language for Special Purposes*, eds I. Simonnæs, Ø. Andersen, and K. Schubert (Berlin: Frank and Timme GmbH), 289–308. Avaialble online at: https://archive-ouverte.unige.ch/unige:135923 (accessed April 26, 2022).

Rainer, F. (2013). Can relational adjectives really express any relation ? *An onomasiological perspective. SKASE J. Theor. Linguist.* 10, 12–40. Available online at: http://www.skase.sk/Volumes/JTL22/

Riloff, E. (1999). "Information extraction as a stepping stone towards story understanding," in *Understanding Language Understanding: Computational Models of Reading*, rds. A. Ram and K. Moorman (Cambridge, MA: The MIT Press), 435–460.

Rojas-Garcia, J., and Cabezas-Garcia, M. (2017). "Use of knowledge patterns for the evaluation of semi automatically-induced semantic clusters," in *New Challenges for Research on Languages for Special Purposes. Selected Proceedings from the 21st LSP-Conference 28-30 June 2017, Vol. 154*, eds I. Simonnæs, Ø. Anderson, and K. Schubert (Bergen, Norway: Forum für Fachsprachen-Forschung), 121–140.

Roller, S., Kiela, D., and Nickel, M. (2018). "Hearst patterns revisited: automatic hypernym detection from large text corpora," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2:*

Short Papers) (Melbourne, VIC: Association for Computational Linguistics), 358–363. doi: 10.18653/v1/P18-2057

Schauble, L. (2016). The development of scientific reasoning in knowledge-rich contexts. *Dev. Psychol.* 32, 102–111. doi: 10.1037/0012-1649.32.1.102

Schumann, A. K. (2011). "A case study of knowledge-rich context extraction in Russian, in *Short papers of the 9th International Conference on Terminology and Artificial Intelligence, TIA* (Paris: LIPN, Université Paris 13), 143–146.

Sinclair, J. (1991). *Corpus, concordance, collocation: Describing English Language*. Oxford: Oxford University Press.

Skuce, D., and Meyer,. I. (1991). "Terminology and knowledge engineering: exploring a symbiotic relationship," in *Proceedings of the 6th International Workshop on Knowledge Acquisition for Knowledge-Based Systems* (Banff, AB).

Temmerman, R. (2000). *Towards New Ways of Terminological Description. The Sociocognitive Approach*. Amsterdam; Philadelphia: John Benjamins. doi: 10.1075/tlrp.3

Wüster, E. (1979). *1985. Introduction to the General Theory of Terminology and Terminological Lexicography*. Quebec, QC: Girsterm. Unpublished manuscript.