Narrowing the Gap Between Termbases and Corpora in Commercial Environments

Kara Warburton
City University of Hong Kong
Tat Chee Ave., Kowloon, Hong Kong
E-mail: kara@termologic.com

Abstract

Terminological resources offer potential to support applications beyond translation, such as controlled authoring and indexing, which are increasingly of interest to commercial enterprises. The ad-hoc semasiological approach adopted by commercial terminographers diverges considerably from methodologies prescribed by conventional theory. The notion of termhood in such production-oriented environments is driven by pragmatic criteria such as frequency and repurposability of the terminological unit. A high degree of correspondence between the commercial corpus and the termbase is desired. Research carried out at the City University of Hong Kong using four IT companies as case studies revealed a large gap between corpora and termbases. Problems in selecting terms and in encoding them properly in termbases account for a significant portion of this gap. A rigorous corpus-based approach to term selection would significantly reduce this gap and improve the effectiveness of commercial termbases. In particular, single-word terms (keywords) identified by comparison to a reference corpus offer great potential for identifying important multi-word terms in this context.

Keywords: terminography, corpora, keywords.

1. Introduction

This article describes PhD research that studied terminological data in commercial corpora and termbases with an aim to identify issues and challenges in commercial terminography[1] particularly with respect to term selection. The research is motivated by a perception that conventional terminology theories and terminographical practices are often not applicable in commercial environments, where terminological resources are leveraged in various applications to reduce costs, improve content quality and retrievability, and increase productivity of content creators and translators. In such production settings, the author's experience suggests that semantic-based notions of termhood are less relevant than the reuse potential of a given linguistic expression across the content production chain. This research is an attempt to define some best practices for term selection in commercial settings based on empirical observations.

2. Commercial language as an LSP

There is a consensus in the literature that terminology is the lexis of languages for special purposes (LSPs) (Cabre, 1999; Picht et al, 1985; Rondeau, 1981; Sager, 1990; Dubuc, 1992; Wright, 1997). Our research first addressed the question of whether the language used in a commercial setting[2] can be considered an LSP. If it does not, then one could claim that it does not contain terminology.

Various scholars provide evidence that commercial language does constitute an LSP (Rondeau, 1981; Rey, 1995; Cabre, 1999; Sager, 1990; Pearson, 1998). Scholars generally agree that an LSP (1) is confined to a subject field,

(2) exhibits a closed set of linguistic properties, (3) is used in a specific communicative context for a specific communicative function, and (4) is consciously acquired. According to these broadly-recognised properties, the language used in most companies does indeed constitute a type of LSP, if, as some scholars maintain, the notion of subject field has broadened from a highly-structured objectivist hierarchy of science and technology to an experientialist delimitation that is context- and application-dependent (Rondeau, 1981; Rey, 1995; Cabré, 1999). We therefore conclude that commercial language contains terminology.

3. Multi-purpose terminological resources

There is also ample evidence in the literature that terminological resources should be developed to be multi-purpose in order to address current and future uses beyond translation, such as controlled authoring, indexing, product classification, and search engine optimisation (Knops and Thurmair, 1993; Sager, 1990; Meyer, 1993; Galinski, 1994; Ibekwe-SanJuan et al, 2007; Buchan, 1993; Cabré, 1999; Strehlow, 2001; Jacquemin, 2001; Nazarenko and El Mekki, 2007; Greenwald, 1994; Wettengel and Van de Weyer, 2001; Ahmad, 2001). It would be shortsighted to believe that commercial enterprises are not interested in such applications as they have significant potential for improving content management.

4. Challenging conventional theory

We maintain that the conventional notions of termhood and methods of terminography prescribed by the mainstream theories of terminology are ill-suited as founding principles and methods for commercial terminography. Terminology

---

[1] The term "terminography" refers to the work involved in managing terminology, and "terminographer" to the person who carries out this work. These terms are analogous to lexicography and lexicographer and have been used by other scholars (e.g. Rey,

1995; Sager, 1994; L'Homme, 2004; Pearson, 1998)

[2] For convenience purposes, we henceforth refer to the language used in a commercial setting as "commercial language"

management in a commercial setting is driven by pragmatic concerns such as supporting writers and translators, strengthening marketing collateral, and developing resources that are multi-purpose. All activities of the terminologist must be economically justified. Lombard (2006) emphasises lexical units that have marketing importance or present translation challenges. The notion of what constitutes a term in a commercial setting shifts from traditional semantic criteria to statistical measures of frequency and contextual conditions, such as the visibility of a term to product users. For commercial applications, a term could simply be defined as any lexical entity that, if managed according to certain suitable methods, can bring benefit to the company. The field of terminology is witnessing a shift in the notion of termhood from a representation of an objectivist, language-independent concept to a contextually-dependent expression fulfilling a given communicative purpose (Temmerman et al, 2010; L'Homme, 2005; Cabré, 1999).

Some ideals of the General Theory such as univocity and the onomasiological approach have already been challenged in theories that emphasise communicative, cognitive, and textual aspects of terminology (Temmerman, 1997 and 2000; Cabré, 1999; Rey, 1995; Sager, 1990 and 2001; L'Homme, 2004 and 2005). These challenges resonate well for commercial terminologists. For example, terminology work in production-oriented settings is almost always semasiological (Bowker, 2003; Martin and van der Vliet, 2003). An ad-hoc, rather than thematic, approach is the norm (Wright et al, 1997). And terminological variation is common in many text types (Jacquemin, 2001; Sager, 1990; Rogers, 1997). Commercial texts are no exception. We join these scholars in calling for pragmatisation of the methodologies of modern terminography.

5. Correlating termbases and corpora

To be effective resources for the various applications that may need them, company termbases need to reflect the corporate language in use, and at the same time guide corporate language towards improved use.

Reflecting language in use means that there needs to be a high degree of correspondence between the company termbase and the company corpus, reflecting a descriptive approach to terminology. Using four companies as case studies, we provide empirical evidence of a significant gap between the two, we propose various linguistic explanations for this gap, and we explore ways to reduce the gap. This was the main focus of our research.

On the other hand, guiding language towards improved use means that the termbase needs to include some terms that may be absent or infrequent in the corpus. Three of the companies in our research are engaged in some level of controlled authoring, which requires a prescriptive approach. However, only one of the four companies

actually includes terms required for controlled authoring in the termbase. Our analysis of the gap needs to take this into consideration. It should also be noted that all the companies use computer-assisted translation (CAT) tools.

6. Methodology

We obtained a termbase and a corpus from each company with the precondition that there be an optimal level of correspondence between the two. The termbases ranged in size from about 2,300 to 6,700 entries and varied considerably in structure. The termbases were then converted to TBX and imported into TermWeb [3], a terminology management system (TMS), so that they could be compared and manipulated for the research. The data model and data categories were compared to identify common features and potential problems

Our research was confined to English terms. Certain filters had to be applied to the termbase terms to ensure that we were working with comparable data between all four companies and to provide an effective measure of the problematic gap between terrmbases and corpora. Not all the terms in a termbase can necessarily be expected to occur in the corpus in any significant degree. In particular, terms that are in the termbase for purposes of *reducing* or *preventing* usage should not occur in the corpus, assuming that users are adhering to those guidelines (terms with a usage indicator of "deprecated," for instance). Any gap between the termbase and the corpus that is attributed to such terms is actually justified. We undertook to eliminate such terms from those considered in the gap measurement.

Another concern were lexical units in the termbase that were deemed to belong to the general lexicon, such as "do," "present," and "person." One company termbase contained over 600 such expressions (most of which were marked and could therefore be isolated automatically), while the others had very few. Some were even function words such as pronouns and prepositions. The presence of such units can be explained by the fact that the company in question uses its termbase for controlled authoring, in addition to translation. Nevertheless, the number of concordances of these units in the corpus would be very high, thus skewing the gap measure for the so-called "real" terms. Our experience also suggests that including them would exceed the processing capabilities of the batch concordancer. (Batch concordances of the termbase terms *without* the general lexicon units took several weeks and had to be carried out on smaller portions of the termbase terms at a time and the results then recompiled.)

We applied several such selection procedures in order to produce a set of termbase terms that was comparable between the four companies and that could reasonably be expected to occur in the corpus. We called this set of terms "corpus-valid" for the purposes of our research.

The corpora came in various proprietary formats and required significant pre-treatment for this research. They ranged in size from 4 million to 20 million words. A normalisation factor of 4 million was applied for comparison purposes.

After generating a plain list of corpus-valid termbase terms by using a TermWeb export, we used the WordSmith batch condordance function to measure the frequency of occurrence of the termbase terms in the corpus. This produced a baseline figure against which we could measure various gap-reducing techniques. The baseline figure revealed a significant gap between the termbases and the corpora. We then conducted interviews with the corporate terminologists to determine if there were any explanations for the gap and to clarify the terminographic methods and uses of terminology in the company. In several cases these interviews resulted in an updated corpus and/or termbase being supplied due to errors or omissions that were discovered in the original shipment, and we had to repeat the batch concordance and generate new baseline figures.

Using the concordance results, we grouped the termbase terms according to four normalised frequency ranges: those that do not occur in the corpus at all (nonextant terms), those that occur infrequently, and those that occur frequently, and identified linguistic properties that were contributing to the gap in each case. For some analyses, we broke these categories down more granularly, such as very infrequent and very frequent.

It is generally acknowledged that terms comprising more than one word (multi-word terms, or MWTs) predominate as terminological units (Meyer and Mackintosh, 1996; Nagao, 1994; Cabre, 1999; Maynard and Ananiadou, 2001; Knops and Thurmair, 1993). We therefore considered the possibility that some important MWTs were missing from the termbase. To find these terms, we used the keyword functionality of Word Smith to identify domain-specific unigram terms, and then ran concordances using these terms as search nodes to determine whether they are productive in forming important MWTs.

7. Results

7.1 Terminographical practices

Through examining the termbases, we discovered that certain terminographical practices necessary for developing multi-purpose resources where neglected, thereby reducing the repurposability and diminishing the value of the termbase. For instance, both variants and non-nouns were under-documented, possibly due to the influence of classical tenets that continue to dominate the literature, i.e. that variants should be avoided and that terms are nouns. Important data categories such as part-of-speech and subject field were missing. Only one company regularly records definitions (and this is because it is used as a source of customer-facing glossaries). Further, there were violations of best practices such as concept-

orientation, term autonomy, and data elementarity. Fields were too frequently mis-used. Several data categories are placed at levels inconsistent with ISO 16642 (TMF). In some cases, these violations were explained by limitations in the TMS, or by business-based driving factors.

For instance, the companies adopt different approaches to translate software strings consistently between the UI and peripheral materials such as online help and documentation: (1) single-sourcing the translations of software UI strings directly from the software files into peripheral materials, (2) leveraging a combination of translation memories and lexical resources in a sequential process, (3) maintaining a separate localisation environment and termbase for translating software strings, and (4) importing software strings and their translations into the termbase. The third and fourth options result in software strings being stored in termbases. And yet, most software strings are not "terms." If not properly marked so that they can be filtered from "real" terms, repurposing the data is severely impacted.

A violation of the principle of data elementarity occurs when two terms are recorded in one instance of the field reserved for terms. Yet this practice was adopted by two of of the companies to address two needs: (1) to prompt writers to introduce two synonymous terms to end-users on first occurrence, such as a full form and an acronym or two synonyms, and (2) to address limitations of the automatic term recognition function of the CAT tool. The following are examples:

```
<term>modular arithmetic (also called clock
arithmetic)</term>
<term>moving average (MA)</term>
```

Only one of the companies handled this situation properly by also including, in the same entry, the two terms separately, such as:

```
<term>modular arithmetic (also called clock
arithmetic)</term>
<term>modular arithmetic</term>
<term>clock arithmetic</term>
```

7.2 The gap

There are two types of terms that contribute to the gap between termbases and corpora: (a) terms in the termbase that are absent or infrequent in the corpus, and (b) important terms in the corpus that are missing in the termbase.

The baseline gap between the termbases and the corpora in our study is significant in all cases; the percentage of corpus-valid termbase terms that do not occur at all (nonextant terms) or occur very infrequently (10 times or less in a corpus of 4 million tokens) ranges from 35 percent to 73 percent. We call these under-performing terms.

We also demonstrated with empirical evidence that frequently-occurring terms, particularly MWTs, are under-

represented in termbases; only three to eight percent of the termbase terms occur very frequently in the corpus. Even for the best performing termbase, only 37 percent of the terms occur with moderate frequency in the corpus.

7.2.1. Causes of the gap

While some of the gap can be attributed to non-linguistic factors, such as files potentially missing from the corpus, we were able to identify some linguistic causes.

The two termbases with the smaller gap presented the following characteristics compared to the termbases with the larger gap:
a) They included significantly more variants (acronyms, abbreviations, spelling variants).
b) They included significantly fewer terms in initial upper case.
c) They included fewer long MWTs, specifically, fewer terms 3 tokens or longer in length.
The termbase with the smallest gap also contained the most non-nouns. We later discovered that 25 percent of the very frequent terms are verbs and adjectives, a figure that is significantly higher than the industry benchmark of only 10 percent (according to which nouns account for approximately 90 percent of termbase terms).

We also demonstrate with empirical evidence that acronyms and other truncated forms of multi-word terms are more frequently used than their longer counterparts.

7.2.2. Setting optimal term boundaries

Since overall, the frequency of a term decreases as length in tokens increases, we investigated whether certain MWTs contain non-essential parts which, if removed, would render the term more productive as an element of the corpus. We investigated a random selection of nonextant terms that contained a word of potentially minor significance, and found that in almost every case a greater matching could be achieved by adjusting the term boundary accordingly. The presence of a non-essential word in a MWT can significantly lower its match rate in the corpus. Setting term boundaries that are optimised to achieve the repurposing demands for production-oriented commercial applications is a challenge for terminologists.

7.3 Keywords

Keywords are domain-specific unigram terms. They are identified by comparing the frequency of unigrams in the studied corpus with their frequency in a reference corpus. Words that are significantly more frequent in the studied corpus are keywords. Keywords have already proven effective in several research projects focussing on term extraction (Drouin, 2003; Chung, 2003; Kit and Liu, 2008) for producing lists of term candidates. Our approach is novel in that it invevstigates the potential of keywords, coupled with various functions of a concordancing software such as DICE relationship measures, to identify the most productive collocates for *human* term selection.

We selected 116 of the top-ranked keywords for further examination. From this set, we wished to identify keywords that are under-represented in the termbase, that is, the termbase is missing important MWTs that contain the keyword. (We refer to terms that are missing from the termbase as "undocumented.") For each keyword, we identified the set of termbase terms that contained that keyword and calculated their total occurrence in the corpus. We then compared that figure to the frequency of occurrence in the corpus of the keyword itself. Keywords that presented the largest gap between these two figures are more likely to lead to the identification of important undocumented MWTs. Based on this calculation, we identified 21 keywords as having the highest potential for discovering undocumented MWTs. We then examined the concordances and collocates of those 21 keywords.

When the occurrence of termbase terms containing a top-ranking key-word is low compared to the occurrence of the keyword alone, we proved that some very important MWTs are indeed missing from the termbase. These MWTs can be easily identified using a concordancing software. Application of the DICE relationship measure will identify additional important MWTs that would otherwise be overlooked by using the raw frequencies alone. We also discovered that when the number of termbase terms containing the keyword is relatively high, and yet there is still a significant gap between their total frequency and the frequency of the keyword alone, there is a high incidence of redundant and under-performing terms in the termbase based on this keyword. Finally, we observed that keywords that are absent or rare in the reference corpus are highly domain-specific. Therefore it is extremely important to document them, and the MWTs that they form, in a company termbase even if they are infrequent in the company corpus.

8. Conclusions

The main objective for managing terminology in a commercial setting is to reduce costs for content authoring and translation and produce terminological resources that are repurposable. In effect, terminology work must account for terms in active use in order to enable productive re-use. This research shows that commercial termbases are not optimised with respect to the terms that they contain, and that term selection should be more corpus-based. However, terminologists are generally more familiar with conventional theory and methodology which tends to underestimate the role of large-scale corpora in term identification. Terminologists working in companies and other production-oriented settings rarely use large-scale corpora, NLP technologies such as concordancers, or even term extraction tools. We suggest that the use of such resources would significantly improve the value and repurposability of commercial termbases.

This research leads us to also reflect on the theoretical and methodological framework for commercial terminography. An *application-oriented* terminology theory and

methodology is needed, one that takes into consideration the pragmatic concerns of commercial terminologists.

9. Limitations and future work

Termbases in general are not specifically developed to be representative of any particular well-delineated corpus. Furthermore, obtaining a corpus from a company is very difficult. Although we made a special effort to prevent this, it is possible that the corpora were lacking a few files. An incomplete corpus could explain some of the observed gap between the termbases and corpora. Nevertheless, the observations with respect to the types of contributing linguistic factors remain valid even though we cannot be precise as to the scope of their contributions.

The findings of a study limited to four IT companies cannot necessarily be generalised across the entire commercial sector. We therefore view this research as a spring-board for further research and validation. Finally, the use of a corpus that could be part-of-speech tagged, as well as termbase entries that consistently included the part-of-speech data category, could lead to additional findings about the morphosyntactic properties of productive and un-productive terms.

As we noted, using keywords has already been investigated for automated term extraction. It would be interesting to compare the performance of both approaches, i.e. (a) keyword-based automated extraction of MWTs followed by manual cleanup of the term candidates, and (b) keyword-based human identification of MWTs using concordancing tools and relationship measures.

With respect to defining a theoretical and methodological framework for commercial terminography, a promising direction may be to consider the potential contributions of LSP lexicography, and also, to pursue and further develop the notion of "textual terminology" proposed by Bourigault and Slodzian 15 years ago.

10. Acknowledgements

11. References

Ahmad, K. (2001). The Role of Specialist Terminology in Artificial Intelligence and Knowledge Acquisition. In S. Wright & G. Budin (Eds.), *Handbook of Terminology Management*, Vol. 2. Amsterdam: John Benjamins Publishing Company, pp. 809--844.

Bourigault, D. & Slodzian, M (1999). Pour une terminologie textuelle. *Terminologies nouvelles*, V. 19, pp. 29--32.

Bowker, L. (2003). Specialized Lexicography and Specialized Dictionaries. In P. van Sterkenburg (Ed.), *A Practical Guide to Lexicography*. Amsterdam: John Benjamins Publishing Company, pp. 154--164.

Buchan, R. (1993). Quality Indexing with Computer-aided Lexicography. In H. Sonneveld & K. Loening (Eds.), *Terminology: Applications in interdisciplinary communication*. Amsterdam: John Benjamins Publishing Company.

Cabré, M. T. (1999). *Terminology – Theory, methods and applications*. Amsterdam: John Benjamins Publishing Company.

Chung, T. (2003). A Corpus Comparison Approach for Terminology Extraction. *Terminology*, 9(9), pp. 221--245.

Drouin, P. (2003). Term Extraction using Non-Technical Corpora as a Point of Leverage. *Terminology*, 9(1), pp. 99--115.

Dubuc, R. (1992). *Manuel Pratique de Terminologie*. Quebec: Linguatech.

Galinski, C. (1994). Exchange of Standardized Terminologies within the Framework of the Standardized Terminology Exchange Network. In S. Wright & R. Strehlow (Eds.), *Standardizing and Harmonizing Terminology : Theory and Practice*. Philadelphia: American Society for Testing and Materials, pp. 141--149.

Greenwald, S. (1994). A Construction Industry Terminology Database Developed for Use with a Periodicals Index. In S. Wright & R. Strehlow (Eds.), *Standardizing and Harmonizing Terminology : Theory and Practice*. Philadelphia: American Society for Testing and Materials, pp. 115--125.

Ibekwe-SanJuan, F.; Condamines, A. & Cabré Castellvi, M. (2007). *Application-Driven Terminology Engineering*. Amsterdam: John Benjamins Publishing Company.

Jacquemin, C. (2001). *Spotting and Discovering Terms through Natural Language Processing*. Cambridge: The MIT Press.

Kit, C. and Liu, X. (2008). Measuring Mono-word Termhood by Rank Difference via Corpus Comparison. *Terminology*, 14(2), pp. 204--229.

Knops, E. and Thurmair, G. (1993). Design of a Multifunctional Lexicon. In H. Sonneveld & K. Loening (Eds.), *Terminology: Applications in Interdisciplinary Communication*. Amsterdam: John Benjamins Publishing Company, pp. 87--109.

L'Homme, M. (2004). *La Terminologie: Principes et Techniques*. Montreal : Les Presses de l'Université de Montréal.

L'Homme, M. (2005). Sur la notion de "terme." *Meta*, 50(4), pp. 1112--1132.

Lombard, R. (2006). Managing Source Language Terminology. In K.J. Dunne (Ed.), *Perspectives on Localization*. Amsterdam : John Benjamins Publishing Company, pp. 155--171.

Martin, W. & Van der Vliet, H. Design and production of terminological dictionaries. In P. Van Sterkenburg (Ed.), *A Practical Guide to Lexicography*. Amsterdam : John Benjamins Publishing Company, pp. 333--349.

Maynard, D. & Ananiadou, S. (2001). Term Extraction Using a Similarity-based Approach. In D. Bourigault, C. Jacquemin & M.-C. L'Homme (Eds.), *Recent Advances in Computational Terminology*. Amsterdam: John

Benjamins Publishing Company, pp. 261--278.

Meyer, I. (1993). Concept Management for Terminology: A Knowledge Engineering Approach. In S. Wright & R. Strehlow (Eds.), *Standardizing Terminology for Better Communication: Practice, Applied Theory, and Results*. Philadelphia: American Society for Testing and Materials, pp. 140--151.

Meyer, I. & Mackintosh, K. (1996). The Corpus from a Terminographer's Viewpoint. *International Journal of Corpus Linguistics*, 1(2), pp. 257--285.

Nagao, M. (1994). A Methodology for the Construction of a Terminology Dictionary. In B. Atkins and A. Zampolli (Eds.), *Computational Approaches to the Lexicon*. Oxford: Oxford University Press, pp. 397--411.

Nazarenko, A. & El Mekki, T. (2007). Building Back-of the-book Indexes? In F. Ibekwe-San Juan, A. Condamines & M. Cabré Castellvi (Eds.), *Application-Driven Terminology Engineering*. Amsterdam: John Benjamins Publishing Company.

Pearson, J. (1998). *Terms in Context – Studies in Corpus Linguistics*. Amsterdam: John Benjamins Publishing Company.

Picht, H & Draskau, J. (1985). *Terminology: An Introduction*. Denmark: LSP Centre, Copenhagen Business School.

Rey, A. (1995). *Essays on Terminology*. Amsterdam: John Benjamins Publishing Company.

Rogers, M. (1997). Synonymy and Equivalence in Special-Language Texts. In A. Trosborg (Ed.), *Text Typology and Translation*. Amsterdam: John Benjamins Publishing Company, pp. 217--246.

Rondeau, G. (1981). *Introduction à la Terminologie*. Montreal : Centre éducatif et culturel, Inc.

Sager, J. (1990). *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins Publishing Company.

Sager, J. (1994). What's wrong with "terminology work" and "terminology science"? *Terminology*. Amsterdam: John Benjamin's Publishing Company, 1(2), pp. 375--381.

Sager, J. (2001). Terminology Compilation: Consequences and Aspects of Automation. In S. Wright & G. Budin (Eds.), *Handbook of Terminology Management*, Vol. 2. Amsterdam: John Benjamins Publishing Company, pp. 761--771.

Strehlow, R. (2001). Terminology and Indexing. In S. Wright & G. Budin (Eds.), *Handbook of Terminology Management*, Vol. 2. Amsterdam: John Benjamins Publishing Company, pp. 419--425.

Strehlow, R. (2001). The Role of Terminology in Retrieving Information. In S. Wright & G. Budin (Eds.), *Handbook of Terminology Management*, Vol. 2. Amsterdam: John Benjamins Publishing Company, pp. 426--444.

Temmerman, R. (1997). Questioning the Univocity Ideal. The Difference between Socio-cognitive Terminology and Traditional Terminology. *Hermes, Journal of Linguistics*, 18, pp. 51--90.

Temmerman, R. (2000). *Towards New Ways of Terminology Description: The Socio-cognitive Approach*. Amsterdam: John Benjamins Publishing Company.

Temmerman, R., Kerremans, K., De Baer, P. (2010). Competency-based Job Descriptions and Termontography. The Case of Terminological Variation. In T .Steurs & F. Steurs (Eds.), *Terminology in Everyday Life*. Amsterdam: John Benjamins Publishing Company, pp. 179--191.

Wettengel, T & Van de Weyer, A. (2001). Terminology in Technical Writing. In S. Wright & G. Budin (Eds.), *Handbook of Terminology Management*, Vol. 2. Amsterdam: John Benjamins Publishing Company, pp. 445--466.

Wright, S. (1997). Term Selection: The Initial Phase of Terminology Management. In S. Wright & G. Budin (Eds.), *Handbook of Terminology Management*, Vol. 1. Amsterdam: John Benjamins Publishing Company, pp. 13--23.

Wright, S. & Wright, L. (1997). Terminology Management for Technical Translation. In S. Wright & G. Budin (Eds.), *Handbook of Terminology Management*, Vol. 1. Amsterdam: John Benjamins Publishing Company, pp. 147--159.