Localization and terminometrics:
Measuring the impact of user involvement on terminology

Barış Bilgen

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Translation Studies

School of Translation and Interpretation
Faculty of Arts
University of Ottawa

Table of Contents

List of Tables

List of Figures

List of Acronyms

AoIR:	Association of Internet Researchers
ATA:	American Translators Association
GDT:	Grand dictionnaire terminologique
FOSS:	Free Open-Source Software
IC:	Implantation Coefficient
LISA:	Localization Industry Standards Association
MS:	Microsoft
UGT:	User-Generated Translation
UNDP:	United Nations Development Programme

Acknowledgments

I would like to thank my supervisor Jean Quirion for his extraordinary generosity and genuine support on all levels. I would also like to extend my gratitude to my examiners for their feedback; and to the University of Ottawa community and my colleagues at the School of Translation and Interpretation for their support. Thanks to my family, friends and loved ones for supporting me throughout my studies. A special thanks to Loon Choir for the music and the adventures, and for keeping me sane and in Ottawa.

Thanks to Asst. Prof. Pierre-Jerôme Bergeron of the Department of Mathematics and Statistics of the University of Ottawa for his feedback on statistical matters concerning this project. Thanks to Owen Weiss and Celeste Côté for agreeing to assist in the data analysis process. Thanks to Trish Van Bolderen and a special thanks to Kate "Kiraz" MacFarlane for their assistance in proofreading this dissertation, and for their invaluable support.

This research was supported by the Social Sciences and Humanities Research Council of Canada and the Fonds de recherche du Québec - Société et Culture.

Fonds de recherche Société et culture Québec

Social Sciences and Humanities Research Council of Canada | Conseil de recherches en sciences humaines du Canada | Canada

Abstract

Online collaborative translation has received increased attention from Translation Studies, mostly with a focus on explaining the various models it exhibits and the factors that shape these models. This study takes a new perspective on this phenomenon by focusing on its outcome through the lens of terminology. A terminometric analysis is carried out on the terminology used in the discussions of Francophone users on online forums of the Ubuntu-Québec open-source software community. The implantation of terms used in the forums and those stored in a selection of major term banks is examined with the objective of identifying potential correlations between term implantation and term formation patterns. The examination indicates that most terms formed through the use and modification of existing linguistic resources have higher implantation rates than terms formed through the creation of new lexical items. A new avenue of terminometric research is introduced by shifting the focus from the institution to the community, aligning with the global shift in content production and distribution. The study provides insight into online collaboration in the context of localization and points out correlations between term formation patterns and term implantation. These observations can mark a starting point for terminological decision-making that is informed by user behaviour and may thus improve the reception of localized content by adapting to users' terminological expectations.

Résumé

La traduction collaborative en ligne attire de plus en plus l'attention de la traductologie, qui s'intéresse notamment à l'explication de ses différents modèles et les facteurs qui les démarquent. La présente étude adopte une nouvelle perspective sur ce phénomène en se concentrant sur ses résultats dans le domaine de la terminologie. La terminologie utilisée dans les discussions des utilisateurs francophones sur les forums de la communauté du logiciel libre Ubuntu-Québec est soumise à une analyse terminométrique. Les taux d'implantation des termes utilisés dans les forums et ceux stockés dans une sélection de grandes banques de données terminologiques sont examinés afin d'identifier des corrélations potentielles entre l'implantation terminologique et les modèles de formation terminologique. L'analyse indique que la majorité des termes formés par l'utilisation et la modification des ressources linguistiques existantes ont des taux d'implantation plus élevés par rapport à des termes formés à travers la création de nouveaux éléments lexicaux. Une nouvelle voie de recherche en terminométrie est ouverte en déplaçant l'accent mis sur l'institution vers la communauté, s'alignant ainsi avec la tendance globale dans la production et la distribution de contenu. L'étude donne un aperçu de la collaboration en ligne dans le contexte de la localisation et présente des observations sur la relation entre les modèles de formation terminologique et l'implantation terminologique. Ces observations peuvent marquer un point de départ pour la prise de décision terminologique dans le contexte de la traduction collaborative en ligne, qui serait informée par le comportement des utilisateurs et, donc, qui pourrait ainsi améliorer la réception du contenu localisé en répondant aux attentes terminologiques des utilisateurs.

1. INTRODUCTION

This chapter describes the context and motivation for this thesis (1.1), sets out the research questions to which it aims to respond, defines the hypothesis, outlines the objectives it aims to achieve (1.2), and establishes its scope and limitations (1.3).

1.1. Context and Motivation

Software *localization*[1] serves as a means of importing new technology, which is likely to come with new *terminology*[2] to express new concepts. Throughout the process of localization, as the software product and its various components are adapted for the target locale, linguistic and terminological choices are made by individuals or institutions participating in the process of localization in order to accommodate new technological concepts by finding terminological solutions. Such choices may be shaped by many factors such as professional (e.g. corporate/client guidelines) or ideological (e.g. language planning efforts to preserve linguistic identity) concerns. The outcome of these choices shapes domain terminology in the target language, especially when the domain is new and its terminology is in the process of being formed. This study constitutes a focused attempt to reach a better understanding of the outcome of terminological choices made in software localization.

Software localization emerged in the 1980s as software companies aimed to reach wider international customer bases. As computer use increased across the globe, it generated a diverse

[1] Esselink defines *localization* as "the translation and adaptation of a software or web product, which includes the software application itself and all related product documentation," for a specific target locale, representing "a specific combination of language, region, and character encoding" (Esselink 2000: 1).
[2] *Terminology* is used here to refer to terms of a specialized domain. Sager (1990: 57) defines a term as a linguistic representation of a concept in specialized language which is based on systematized rules or principles of designation.

user[3] profile, linguistically and otherwise. In its initial applications, localization was a last-minute detail in the production cycle. The software product would be designed, developed and marketed in the main language, most commonly English, and later localized based on local market needs and potential. However, this type of separation between production and localization created various difficulties and *internationalization*[4] and localization later became integral parts of the production cycle (Esselink 2006: 22, 23).

Increasing competition and time constraints led to the simultaneous shipment of multiple language versions of a software product. Accordingly, the development process had to be reshaped so that localization could progress simultaneously with its other components. Companies that hire language service providers or in-house translators, or those that engage volunteers for localization purposes release multiple, if not all, language versions of their software applications simultaneously in an effort to reach the widest customer base as quickly as possible. Localization projects are not entirely finalized and shelved once the software is released, with updates and additions being released almost around the clock for the duration of the product life cycle – in other words, as long as the product is maintained and supported by the software publisher. This implies that localization is an integral part not only of the production cycle, but also of product support. Moreover, localization is no longer limited to software localization but now concerns all types of digital *content*[5] such as websites and mobile applications (Esselink 2006: 28). This means increased workload and demand as well as tighter

[3] In the context of this project, the term *user* refers to an individual who uses computers and computer applications (including websites and all types of software).
[4] *Internationalization* refers to measures taken in software development to provide a level of linguistic and cultural neutrality to the fundamental structure of the software in order to allow for the modifications required for localization, and to avoid subsequent modifications on the original program for any specific locale (Esselink 2000:2; Schmitz 2014: 451).
[5] Within the context of this project, the term *content* is used to refer to any type of material or service published or provided online, including but not limited to image, text, video, audio and software.

schedules for localization service providers.

The need for speed in the competitive information technology market has been driving new localization models, most significantly in the form of increasing user input in the localization process. Depending on the level of user involvement, this may be in the form of post-release user feedback, or in the case of higher involvement, it is observed in the form of localization projects undertaken by users. In such cases, users fully or partially assume localization-related tasks and therefore eliminate or reduce the workload on localization professionals, generally accelerating the release of localized content. Recent examples of such initiatives include popular websites and software applications, such as the social networking website Facebook (O'Hagan 2009: 112), and the *open-source*[6] web browser Mozilla Firefox (Beatty and Malolepszy 2013). In these examples, users of Facebook and Mozilla have been engaged through various forms of non-monetary motivation via different input platforms and methods (e.g. Facebook's translation app; Mozilla's localization tool, Pontoon), ultimately for the same objective, i.e. localization for multiple locales without financial compensation.

Various terms have been used to describe this localization model, including "crowdsourcing" (Howe 2006), "massive online collaboration" (Désilets 2007), "online volunteer translation" (Bey et al. 2008) and "user-generated translation" (O'Hagan 2009). Each term carries some level of nuance. For instance, *crowdsourcing* originates in the term *outsourcing,* which Merriam-Webster's online dictionary defines as "to procure (as some goods or services needed by a business or organization) under contract with an outside supplier"[7]. This definition, which is based on the activity of procurement, reflects a business management perspective. In

[6] The term *open* refers to unrestricted access by users of a software product or website to facilities that enable the modification of its various components. The term *open source* (*open-source* when used as adjective) is used to refer specifically to a software product which has a source code that is open for modification.
[7] http://www.merriam-webster.com/dictionary/outsourcing, last accessed 8 August 2016.

comparison, the terms *massive online collaboration, online volunteer translation* and *user-generated translation* describe the concept from a broader perspective.

The localization model described by the terms mentioned above is essentially based on the contributions of users communicating online, who provide their services under the impetus of various types of motivational factors such as personal emotional investment due to physical proximity to benefits provided by the service (cf. Munro 2010) or financial compensation at much lower levels than the cost of hiring professionals (e.g. Amazon's Mechanical Turk, cf. Bloodgood and Callison-Burch 2010). In a way, this model taps into the unique capacity of users in their knowledge, experience and expectations from the software, in order to make it accessible to other users. Dray and Siegel (2006) discuss the connection between localization and user-centred design and highlight the significance of user concerns and the role of users' familiarity with the context of software use. They emphasize how user input needs to be incorporated in the early stages of software development to fulfill the needs and expectations of software users. These observations highlight the benefits drawn from users working for users, which is a phenomenon currently shaping the virtual world.

The emergence and widespread use of interactive online platforms have changed the dynamics of content production and distribution irreversibly. On social networking and content sharing web platforms, Internet users now freely produce and distribute various forms of content. The Internet has provided online communities of different scales with the technical potential to undertake large-scale initiatives and reach wide audiences on a similar level to governments and institutions (cf. Shirky 2008) and thus with a comparable level of potential influence on language. This project represents an acknowledgment of the increasing role played by global communities in the production and distribution of content (including but not limited to linguistic

content), which marks a shift from institutional control on these matters. Simply put, this shift needs to be reflected in the study of language.

With the motivation to gain a better understanding of the terminological outcome of collaborative[8] localization, this study examines the terminology that is used to refer to concepts in collaboratively localized software on the basis of term implantation[9] rates. For this purpose, terminometric research [10] methodology presents an opportunity to reach objective and reproducible results representing term implantation. Although the focus of conventional terminometric research has been on the evaluation of language planning efforts through institutional communications, this doctoral thesis project shifts the focus of the methodological approach from the institution to the community of users of a localized software product.

As a field that is undergoing rapid change, localization presents immense potential for exploration. The observation and description of localization practices and norms represent a major portion of recent localization-related research. Yet, the reception of localized products, especially those localized with the use of emerging localization models, remains a topic to be explored in detail. In particular, there is little to no scientific research on the outcome of user involvement in this context to respond to the multitude of questions that may arise. For instance, is there a difference in the outcome of localization between new collaborative localization models and more conventional ones? Do localization models have any influence on target language terminology? The objectives of this study, which are outlined below, reflect the motivation to address some of these questions.

---

[8] Within the context of this study, the term *collaborative* is used to designate platforms or initiatives involving the participation of multiple individuals in a web-based setting with a shared goal, which can range from sharing content to undertaking specific tasks in a shared manner.

[9] According to Auger (1986: 51) *implantation* of terminology refers to the adoption of terms by users of language to express specialized concepts, by using them actively in authentic communicative instances.

[10] Quirion (2003a: 30) defines *terminometric research* as research that is "designed to measure the degree of implantation of all designations referring to a single concept or to a set of concepts."

1.2. Research Question, Hypothesis and Objectives

In light of the context and motivation presented above, the research question that guides this study can be formulated as follows: Can the examination of terminology in users' discussions regarding localized software reveal patterns in the formation of software terminology that can be replicated or avoided to foster term implantation?

The terminology used in the interface of a collaboratively localized software product is the result of terminological choices made by the community of users involved in the process of localization. Thus, terminology used in online discussion forums by a larger group of users of the same software product (including those who may be involved in the collaboration in various capacities, as well as those who are not actively involved) can serve as evidence of terminological choices made in such collaborative settings. These choices may carry indicators of users' preferences regarding software terminology. Researchers have noted that software terminology and its standardization are important factors that shape users' experience with localized software (Corbolante and Irmler 1997: 534; Lepouras and Weir 1999: 18, 19). Therefore, software terminology that is standardized in light of user preferences can help improve users' experience with localized software.

The hypothesis is that the examination of term implantation in relation to term formation patterns[11] can reveal the patterns that have higher and lower rates of implantation in users' communications. This can provide a valuable contribution to knowledge as the identification of such patterns could help those making terminological decisions meet the terminological expectations of users by making decisions that are more likely to promote term implantation,

[11] *Term formation* refers to the process with which a concept in specialized language is named based on systematized rules or principles of designation (Sager 1990: 57). The term *pattern* is used in reference to Sager's (1990) patterns of term formation.

thus potentially improving the harmonization of terminology in localized products. Considering the role localization plays in the importation of new technology and concepts, informed terminological decisions that facilitate term implantation can go a long way in improving the reception of localized products by enhancing their accessibility linguistically. Furthermore, this study provides a focused contribution to research on user-generated content and offers a new perspective for future research examining the outcome of this phenomenon.

Finally, this thesis introduces a new context for terminometric analysis in a setting outside language planning. This can pave the way for the use of terminometric research in the assessment of terminology use in various contexts in which communities have more control than institutions over terminological choices, reflecting current trends in content production and distribution. Considering the increasingly larger role assumed by communities rather than institutions in the production and distribution of digital content, this shift in focus is essential for reaching a better understanding of new dynamics that operate in the digital age.

1.3. Scope and Limitations

This section outlines the scope and limitations of the project. First and foremost, it is worth noting that working on a technology-related field that evolves rapidly and constantly is one of the major challenges for this study, as any finding or discussion risks being short-lived as a result of new developments. Therefore, establishing a sound scope and clear definitions of key concepts is of utmost importance in order to be able to reach tenable conclusions that, although limited by several factors, may contribute to the body of knowledge on its subject matter.

The scope of this project covers the analysis of term formation patterns observed in the terminology used in discussion forums that constitute the main platform of communication among the French-speaking users of the open-source operating system Ubuntu in Québec

(Ubuntu-Québec forums) in relation to term implantation rates. The terms studied are selected on the basis of concepts found in a sample taken from the corpus gathered from Ubuntu-Québec forums. Terms designating these selected concepts within the corpus as well as those designating the same concepts as recorded in major Canadian terminology resources, in addition to Microsoft's official terminology database were retained for analysis. The latter resource is consulted in order to ensure representation of mainstream software terminology among the terms that are examined, in consideration of its potential influence on open-source software terminology. It should be noted that the focus is essentially on the terminology used by Ubuntu users, and the examination of Microsoft terminology or its overall influence on open-source software terminology is not within the scope of this study.

The analysis is limited by the available material on the Ubuntu-Québec forums, and by the outcome of the random sampling process that is used in selecting the concepts and terms to be studied. The discussion of potential correlations between term formation patterns and term implantation is restricted to the terms that are examined in this study, the software type (open-source) and product (Ubuntu), and the languages that are covered in the analysis (French, with English as the source of borrowed terms), as well as the specific community (Ubuntu-Québec) that participates in the discussion forums.

The focus of this study is on the terminology of desktop operating systems only[12], and the use of the term *operating system* refers to desktop operating systems, unless specified otherwise. The selection of operating systems as software type for this study is due to their widespread use as a vital component of computer systems that is required for the operation of other applications.

[12] The term *desktop operating system* is used to designate operating systems that run on desktop computers as well as portable personal computers as they are generally interchangeable, as opposed to mobile operating systems that run on mobile devices such as smart phones and tablets. (See PC Magazine Encyclopedia entry on *desktop operating system*, http://www.pcmag.com/encyclopedia/term/41170/desktop-operating-system, last accessed 24 June 2016).

With the increasing use of advanced computing functions in a wide variety of devices (e.g. smartphones, tablet computers, game consoles, entertainment devices), operating systems are also increasingly varied in terms of their functionality and use. The operating system at the centre of this study is Ubuntu, since it is one of the most popular among open-source operating systems (DistroWatch n.d.)[13]. In addition, Microsoft's official terminology resources are consulted since Microsoft Windows is currently the most widely used commercial operating system worldwide (StatCounter 2016).

Finally, terminological choices made in *translation*[14] and localization settings may be influenced by a variety of factors. In examining the terminology in users' communications, the scope of this study is limited for the most part to the investigation of term formation patterns that relate to terminological factors, i.e. factors intrinsic to terms (Quirion 2004: 115). Other factors that may influence term implantation (e.g. "socioterminological and procedural", ibid.) are outside the scope of this study[15]. As much as the outcomes of the study over the long term may contribute to the improvement of linguistic communities' ability to express the modern world in their native languages through translation and localization, the social elements involved in the shaping of terminology are outside its scope, as they constitute a distinct path for exploration from a strictly sociolinguistic or socioterminological angle.

---

[13]DistroWatch (www.distrowatch.com, last accessed 24 June 2016) is a website run by open source enthusiasts that provides a variety of information on open-source software programs, as well as tracking popularity, newly released open-source software, and updates for existing software. It has been referred to as "perhaps the most popular of all Linux desktop sites" (Vaughan-Nichols 2014) and "an amazing treasure trove of information about Linux distributions" (Lynch 2014) within the information technology community.

[14] Within the limits of this project, the term *translation* is understood as designating one of the many components of localization, and is used to refer specifically to the translation of text, and only text, as opposed to localization which encompasses textual as well as extra-textual, multimedia elements.

[15] Although it is not part of the main focus of this study, Quirion's (2004: 197) observation regarding the potentially positive effect of user participation in terminological decision-making is put to the test by this study since open source software localization facilitates user participation in all aspects of software development.

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

The review of the literature is presented in three sections corresponding to the fields of research that shape the theoretical framework of this study: localization, user-generated content and terminology.

2.1. Localization

This section summarizes the evolution of software localization and its current status as an industry, discusses common challenges encountered from a translation perspective, and outlines the role of terminology in the localization process.

2.1.1. History and evolution

Localization entails the translation of textual elements of digital content such as text in the user interface, help and support documents, and marketing elements, as well as the adaptation of audio and visual elements for the linguistic and cultural conventions of the target locale. It is unique among activities involving translation due to the fact that it is carried out and experienced in an interactive, digital and multimedia environment (Schäler 2008: 196).

The evolution of localization went through two major stages before it reached its current and continuing stage of modernization: its beginnings in the 1980s and its maturation in the 1990s (cf. Esselink 2006, Schäler 2008). The beginnings were dominated by a significant amount of ad hoc activity. Localization was essentially driven by market demands and depended on the willingness and capacity of software companies to respond to these demands. Two key concepts emerged as some common practices started to appear in the localization industry:

internationalization and *globalization*[16]. The integration of localization into the production process due to increasing demand for localized software required measures that constitute internationalization to be followed in software development in order to accommodate localization for different locales, using different languages with different alphabets, character sets, stylistic and syntactic rules. Globalization, on the other hand, encapsulates a wide variety of activities ranging from marketing to providing local support, with the goal of rendering the product global.

One of the major indications of the maturing industry in the 1990s was the founding of the Localisation Industry Standards Association (LISA). LISA was active from 1990 to 2011 and played an important role in establishing industry guidelines and standards. With growing interest on the part of the academia and other actors involved in technological research, localization tools and processes developed rapidly and localization established itself as an integral part of software development, as well as a major topic of research among language- and technology-related fields.

Digital content and services are now produced and distributed differently than they were in the 1980s and the 1990s, when the localization industry initially settled on most of its practices. Recent localization practices have followed along the process of transformation brought about by the Web 2.0 phenomenon. Garcia and Stevenson (2009: 29) define Web 2.0 as a migration of content from passive access points towards interactive platforms that replace the need to install software applications locally with interactive real-time access to content and services, as is the case for eBay and Google Docs, for example. Several recent publications

[16] In the context of software localization, *globalization* refers to all activities involved in rendering a software product global, as in launching multiple localized versions across the globe (O'Hagan 2005: 1)

discuss the move from stand-alone, desktop-based localization processes towards decentralized, collaborative and web-based processes (cf. Karsch 2006; Bey et al. 2008; Schäler 2008; Exton et al. 2009; Garcia and Stevenson 2009; O'Hagan 2009; Van Genabith 2009; Mesipuu 2012). The entire volume 10 of *Linguistica Antverpiensia* published in 2011 is devoted to the discussion of issues related to online collaborative translation ranging from its expanding popularity (Kelly et al. 2011) and ethics (McDonough Dolmaya 2011) to translation quality assessment (Jimenez-Crespo 2011) and technology awareness on the part of professional translators (Gough 2011). It is evident that interest in the subject among researchers is growing. However, the reception and outcome of such collaborative initiatives have not been sufficiently documented in the few empirical studies to date. This constitutes the essential gap in existing knowledge identified for the purposes of this study.

Today, most digital content accessed by users is generally no longer in the form of a finalized and fixed product. Rather, it is provided through an interactive web-based platform enabling user contributions and modifications. A web search for most accessed websites in the world leads to several sources that list websites such as Facebook, YouTube and Wikipedia in the top ten[17]. What these three websites have in common is that each one runs entirely on user-generated content. O'Hagan describes the role of the user in the context of collaborative translation as follows:

> The user in UGT *[user-generated translation]* (…) is somebody who voluntarily acts as a "remediator" of linguistically inaccessible products and "direct producer" of Translation on the basis of their knowledge of the given language as well as that of particular media content or genre,

[17] http://www.alexa.com/topsites; http://toolbar.netcraft.com/stats/topsites; http://en.wikipedia.org/wiki/List_of_most_popular_websites, last acessed 24 June 2016.

spurred by their substantial interest in the topic. (O'Hagan 2009: 96)

The social networking website Facebook launched localized versions of its interface as a result of volunteer contributions by its users who translated the interface into several languages and voted on each other's translations. This has become a very popular example of the use of collaborative localization. Contrary to the popular suggestion that this was essentially a financially motivated initiative, the most significant advantage of this approach has been argued to be the expertise of Facebook users in understanding and expressing concepts used on Facebook, as opposed to hired professional translators who may not have the same insight (keynote address by Losse 2008 quoted in O'Hagan 2009: 114; Garcia and Stevenson 2009: 30).

This project is partly motivated by the need to reach a better understanding of the phenomenon of user involvement in localization, which is further discussed in section 2.2.

2.1.2. Challenges

According to Schäler (2008: 196), one of the main reasons why localization imposes a distinct set of challenges is its multimodal nature. Text, audio, video, graphics, and other elements in various formats that the end user may view or interact with can be the object of localization. This challenge is often referred to in localization literature, as researchers point out that the objective of localization is to create a user experience in the target locale equivalent to that of the original language users, with an emphasis on experience as a phenomenon that encompasses more than linguistic and textual elements (Corbolante and Irmler 1997: 518; Mangiron and O'Hagan 2006: 14, 15).

Another commonly mentioned challenge in localization relates to context (or lack thereof) when translating (Schäler 2008: 197). In most localization projects, textual content is

extracted from the software interface for translation purposes. Consequently, as translators translate, they may have little or no recourse to the actual visual context in which their translations are going to be used, which creates a variety of problems due to factors ranging from semantic ambiguity to visual space restrictions. Pym (2005) discusses the non-linearity of electronic texts in localization, and questions the effects of the "loss of discursive linearity." He identifies two ways in which localization both creates and benefits from non-linearity: "leveraging", which "concerns the breaking up of text into parts that have to be worked on at different levels"; and "chunking", which refers to texts being "broken down into minor units that can be mixed and matched to form new texts" (ibid.). On a purely practical level, these practices are functional in localization since the text to be localized needs to be separated from the programming languages that form the software, and chunking allows units of text to be programmed to appear in certain places at certain times as the software is used. This approach to text creates a suitable situation for the use of translation and terminology tools as it facilitates the identification of equivalent terms or larger translation units across languages, which can help ensure consistency within a software product, among different versions of the same product or among different software products of the same genre or type and so on. Nonetheless, Pym is not alone in his concerns regarding the effects of language technologies on language itself. For example, Bowker (2006: 179, 180) discusses the loss of context and coherence in such units of texts in chunks and mentions several ways these practices influence the way translators approach language, ranging from following source text structures too closely to the loss of linguistic diversity. In essence, Bowker concludes that it boils down to using the right tools in the right setting, without over-estimating the capacity of language technologies and under-estimating the importance of professionals' expertise in using such technologies.

In order to address the issues arising from the separation of text from its context in localization, starting with the consolidation of the localization industry, user-interface localization tools have been developed (Esselink 2006: 27). These tools aim to provide access to text in context during localization, so that translators working on a project can view and modify text as it appears in the user interface and thus assess context both for semantic accuracy and visual efficiency. Jimenez-Crespo (2013: 49) points to the expansion of the definition of *text* in the context of localization by highlighting that "texts in localization incorporate interactivity and specific sequencing of events that resemble dialogic exchanges […], and therefore, interactivity is an integral part of the text." Klischewski (2015) demonstrates this in his argument in favour of a dedicated video game localization tool which he claims "will have to feature both context information and support for variables" (ibid. 48). Accounts of recent examples of localization platforms and tools such as BeyTrans (Bey et al. 2008) and Pontoon (Beatty and Malolepszy 2013) demonstrate that their main objective is to provide an integrated environment for translators working in localization and thus afford them more independence and flexibility (Bey et al. 2008: 145; Beatty and Malolepszy 2013: 28).

There are also challenges specific to localization from a business perspective, which shape the practices in the industry. The main factor driving business decisions in the context of localization is the demand for simultaneous shipment of several language versions of a software product. The fast pace and the continuous demand created by the cycle of software development, release and updates require that localization be a truly integral part of the product lifecycle (Esselink 2006: 28; Schäler 2008: 200; Van Genabith 2009: 7). This means that localization service providers can be called upon to work on a software product for as long as the product is supported. With such a demand on the work force come questions about where to find and how

to finance this work force. The conventional answers have been an in-house or outsourced localization team. More recently, a crowdsourced model has emerged.

Fluixa (2004) compares the two conventional options (in-house and outsourced localization), and concludes that outsourcing may prove more efficient for larger companies, since localization service providers handle all the strategy and human resources (HR) management necessary for large volumes of work, they are technologically prepared, and they tend to employ translators in their native countries which means that their use of language is likely up-to-date. On the other hand, this option drives localization costs up and the level of cooperation and localizer involvement down, as well as causing communication issues between the software company and the localization service provider. According to Fluixa (ibid.), an in-house localization team seems to be the better choice for a smaller scale localization operation, as it implies lower costs (except for additional HR and technology investments), better communication and coordination among all the teams involved, linguistic testing advantages, quicker and more efficient problem solving, and more efficient management of updates. The downsides of an in-house localization team are that translators living outside their native country may risk losing touch with the current status of their native language, and that a smaller localization team may have difficulty in meeting tight deadlines. Frank's (2013) assessment of in-house versus external translation teams in the localization setting highlights that external language service providers have grown substantially in the recent years. The author relates this to an increasing demand for localization services, the flexibility offered by external providers in terms of the availability of a variety of services whenever needed, and the management of financial costs. Nonetheless, Frank (ibid.) notes the key advantage of in-house translators as their capacity to work in close proximity with other teams involved in localization such as design and

marketing, which has a significant influence on the quality of localization.

Collaborative localization was brought to the forefront of the translation community after the popular social networking website Facebook crowdsourced its entire localization operation, and another social and professional networking website, LinkedIn, attempted to follow suit. This resulted in an outcry among translators, even leading the American Translators Association (ATA) to declare "crowdsourcing one of the two top threats to the profession and to the association, at the same level as the other top threat they identified, the economic downturn" (Baer 2010). Despite this reaction, both the industry and the research community have shown growing positive interest in the potential for using the energy and the motivation of masses of computer and Internet users to complete large and otherwise costly tasks (Souphavanh and Karoonboonyanan 2005; Eagle 2009; Munro 2010). User-generated content is discussed in further detail in section 2.2, with examples that highlight the advantages and drawbacks of user involvement with a focus on translation and localization settings.

2.1.3. The role of terminology

Different types of digital content, such as software or websites or different sets of functions or services offered through such platforms, often come with or lead to the creation of specialized terminology representing new concepts that appear with the rapid development of computer technology. For example, just a decade ago if someone were to talk about reading a *tweet,* it would have made no sense[18]. Now, standardized terms to designate this new concept in both of Canada's official languages can be found in the Canadian government's official terminology

[18] The social media platform Twitter was launched in March 2006, with the first tweet by co-founder Jack Dorsey: "just setting up my twttr", https://about.twitter.com/company/press/milestones, last accessed 24 June 2016.

database, TERMIUM[19].

The quality of terminology in a software interface directly influences the experience of the user, since it makes up a significant portion of the interface, in addition to its role in user manuals, support documents, and so on. Ensuring concision and consistency in the terminology of a software product or website is a top priority for translators working in localization. Consistency plays a crucial role in terminology harmonization and ultimately standardization[20] (Corbolante and Irmler 1997: 534), which can facilitate comprehension and enhance user-friendliness. Lepouras and Weir (1999: 18, 19) emphasize the importance of the quality of localization for user-friendliness, and point out that in some cases localization may create difficulties that do not exist in the non-localized version, therefore proving to be counter-productive as far as ease of use and user experience are concerned. According to Schmitz:

> "... Terminology is the primary means of communication and knowledge transfer between software developers and end-users via the user assistance material. Consequently, avoiding indeterminate, incorrect and inconsistent use of terms and icons must be one of the major goals of software development, quality assurance and usability testing." (Schmitz 2014: 453)

The priority in translating the terminology in a software product or website aligns with the main objective of localization – to create a user experience in the target locale equivalent to that of the original language users. According to Corbolante and Irmler (1997: 518), "it is essential to try to

---

[19] *tweet / gazouillis,* http://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng&srchtxt=tweet, last accessed 24 June 2016.

[20] In the context of terminology, *standardization* refers to the ideal of establishing a single term designating each specific concept, essentially based on the traditional theory of terminology which aims "to eliminate ambiguity from technical languages (…) in order to make them efficient tools of communication" (Cabré Castellvi 2003:165).

define who the implied users of the localized product will be, i.e. the hypothetical ideal users of a program, and reflect on their expectations and background so as to configure tailor-made solutions." The authors further argue that users' level of experience can be an important determinant for making translation decisions (ibid. 519). This demonstrates how central the user is for the localization process. Several researchers point to the advantages of actual users of software or websites being actively involved in the localization process, as they possess the kind of domain knowledge and experience that is not learned through any academic or professional training, but gained through the user perspective (Dray and Siegel 2006: 294; O'Hagan 2009: 96; Van Genabith 2009: 6).

One of the characteristic properties of terminology work in the localization setting is that a significant amount of effort is required for the formation and translation of new terminology in order to accommodate new concepts that appear with the development of new technology. Corbolante and Irmler (1997: 516) observe that, in the localization context, "terminologists and translators must coin terms more frequently than ordinarily in common other, more stable domains." This implies frequent recourse to various term formation mechanisms, explained further in sub-section 2.3.2.

2.2. User-generated content

This section provides an overview of the variety of practices concerning the production and distribution of content by users of software and web platforms following the emergence of Web 2.0 with a focus on its instances related to translation, localization and terminology, and its implications for research.

2.2.1. Web 2.0 and user-generated content

As indicated in the Introduction (p.3), many different labels have been used to describe instances of large-scale online collaboration. Howe (2006) coined the popular term "crowdsourcing", influenced by the outsourcing model which led corporations to ship work to countries where labour is cheaper. Désilets (2007) uses the term "massive online collaboration" with a broader focus to find a common ground for various instances of similar collaborative initiatives and infer their impact on the world of translation. Haythornthwaite (2009) refers to "peer production" in her examination of patterns in collaborative contexts. Her "light and heavyweight models of peer production" (ibid.) are discussed in the next sub-section (2.2.2). No matter what label is used to refer to this emerging phenomenon, it has generated a wide range of reactions from stakeholders and created new avenues for research in several disciplines.

In the early days of widespread Internet use in the 1990s, websites were mostly static, and users' experience with web content was mainly passive. It was limited for the most part to accessing online content to view or download it, without much room for modification or contribution. Second generation web technologies (as in Web 2.0), however, represent an environment "where people exchange ideas, collaborate and share with others their own user-generated content, thus leading to a web of social interactions that generate business" (O'Hagan 2011: 12).

With technology providing people with the means for instant mass communication, the dynamics of content production and distribution have irreversibly changed. Whereas large institutional entities such as governments, private organizations, or publishers had significant control and power over what was produced and distributed and how, the Internet and especially the Web 2.0 phenomenon have given individuals and communities a significant share of that

control and power, which allows them to reach massive audiences instantly. Technically, there is no difference in the size of the potential audience for any content created or shared by any individual (e.g. an update on social media shared by an individual witnessing an important event) and the same type of content created or shared by an institution (e.g. a news report on the same event by a large news organization). There have been numerous examples of various types of content created or distributed by individuals such as viral videos (cf. Guadagno et al. 2013) reaching massive global audiences through social networks. In addition to enhancing individuals' capacity to reach audiences, changes in technology have also facilitated individuals' capacity to form communities by removing various obstacles. For example, the collaborative art platform HITRECORD[21], started by actor and film-maker Joseph Gordon-Levitt, brings together creators of various art forms ranging from poetry to cinematography from across the globe, forming a unique artistic medium enriched by an exceptional diversity of artist profiles. In addition to facilitating collaboration across physical distances, collaborative platforms also enhance the capacity of local communities to organize in the face of restrictive political and legal establishments, as recently seen during the Arab Spring and Occupy movements (cf. Gerbaudo 2012, Castells 2015).

Surowiecki (2004) discusses the immense potential of online communities and the "wisdom of crowds" effect, emphasizing the biggest strength of the crowd as the diversity of the individuals forming it. This idea is echoed commonly in relevant literature. As Howe (2006) observes, "the most efficient networks are those that link the broadest range of information, knowledge and experience" creating an immense "diversity of intellectual background." A common argument against this approach is its potential for a high rate of noise, which can be due

[21] http://www.hitrecord.org/, last accessed 19 August 2016.

to the massive amount of information contributed by a virtually limitless crowd, and sometimes created by unqualified contributors[22]. However, many have observed that a massively wide base of contributors in most cases makes up for the noise created by those unwanted contributions (Surowiecki 2004:10; Howe 2006; Shirky 2008; Needleman 2011).

Web 2.0 and user-generated content have garnered large amounts of positive and negative reaction. The translation community is a prime example of how the phenomenon is affecting professional communities. In the previously mentioned account by Baer (2010) of the strong reaction by the American Translators' Association (ATA), the key concerns of the community identified by the author are quality, the possibility of losing work, and ethical concerns regarding companies profiting from free labour. O'Hagan (2011) also discusses the outcry from the translation community following crowdsourced translation efforts by popular social networking websites Facebook and LinkedIn, and elaborates on the ethical issues surrounding these websites' profits. Importantly, O'Hagan (2009: 112, 3) highlights the fact that Facebook accounts for the motivation as "primarily not a matter of cost saving. It *[Facebook]* points out that most savings which may have been made on translation in fact have been spent on developing its technology platform with the translation application," and that domain knowledge was the most valuable contribution of the crowd. McDonough Dolmaya (2011) concludes in her study on translation ethics in collaborative contexts that the most significant ethical concern in this type of setting is based on whether free labour is used by corporations for profit. However, she also points out the benefits of the potential outreach provided by such initiatives especially for minority languages. Open collaboration fosters user interest and motivation rather than market size and profit, facilitating wider access to localized content (ibid. 105). With open collaboration, digital content

[22] The term *contributor* is used to refer to users who contribute to a collaborative initiative in any way.

can become accessible to linguistic communities without the need for financial investment. For instance, a study exploring collaborative lexicography and comparing it to conventional lexicography through the example of Wiktionary presents several strengths in this resource ranging from its functionality to its coverage (Meyer and Gurevych 2002: 290). The authors highlight the value and permanence of open collaboration, and conclude that "collaborative lexicography will not replace traditional lexicographic theories, but will provide a different viewpoint that can improve and contribute to the lexicography of the future. Thus, Wiktionary is a rival to expert-built lexicons—no more, no less" (ibid. 291).

There are several instances that serve as evidence that initiatives involving user-generated content produce ground-breaking results in some cases, including emergency situations, by providing solutions to problems that would have otherwise been very costly, both financially and time-wise. For example, Ushahidi[23], which is an online platform launched in Kenya as a crowd-mapping initiative in 2008 to document and map incidents surrounding the conflict following national elections, facilitated resource allocation in response to the conflict in addition to raising global awareness. Eagle (2009) also provides an account of crowdsourced translation and data entry in Kenya, where the goal was to localize mobile phone services for over 60 languages. This would have otherwise been impossible due to financial and time limitations, and thereby would have prohibited access to these services for a significant portion of the population. Likewise, Munro (2010) explains how online collaboration was used following the 2009 earthquake in Haiti for translating messages to emergency services and for providing data on the location of those who needed emergency services so they could be located more quickly and easily. These cases present success stories of initiatives that made a noticeable difference in people's lives, and

[23] https://www.ushahidi.com/about, last accessed 24 June 2016.

such achievements would not have been possible without the use of user-generated content. That being said, they were not without problems. Both Eagle (2009) and Munro (2010) acknowledge the lack of quality assurance, which was mainly due to the urgent nature of services provided; however, both state that the amount of high quality and high speed contributions more than compensated for collaborative input of low quality.

Another success story involving the use of user-generated content that has been around for a relatively longer time is open-source software development. Open-source software is free to use, modify and redistribute, and it facilitates and encourages revisions by large numbers of developers and users (Souphavanh and Karoonboonyanan 2005: v). These are qualities highlighted in the foreword to a study sanctioned by the United Nations Development Programme (UNDP) to evaluate open-source software localization for the public sector. Richard Stallman, who wrote the foreword, is a leading name in the free and open-source software community and one of the leaders of the GNU project[24] which is the foundation of the popular open-source operating system known as Linux[25] (Raymond 1999: 18, 19). The same UNDP study also points out that "a growing number of European national and local administrations have developed or are developing policies to promote the use of FOSS *[free open-source software]* instead of proprietary systems and tools" and that "this not only gives them independence from commercial vendors, but also nurtures their own software development industries" (Souphavanh and Karoonboonyanan 2005: 3). Some of the key advantages of FOSS

[24] The GNU project started in 1984 and led to the development of an open-source operating system with a text interface, which has since been used in various forms as the foundation of several open-source operating systems with graphic interfaces. The name is a recursive acronym for "GNU's Not Unix", in reference to UNIX, which is one of the first ever operating systems created in the 1960s.
(http://www.gnu.org/home.en.html; http://www.unix.org/what_is_unix/history_timeline.html, last accessed 1 July 2016.)

[25] Linux is the basis of the operating system Ubuntu, which is the subject of forum discussions analyzed in this project.

localization mentioned in the study are reduced reliance on imports, independence from the dominance of English as the language of technology, and investment in local software development skills (ibid.).

Open-source software lends itself perfectly to collaborative localization, since it can be freely modified to suit the linguistic and cultural norms of any target locale. Open-source software localization is central to this study as its main research question relates to the outcome of user involvement in the localization process, and its methodology involves the application of terminometric analysis to terms used in the discussion of matters related to a localized open-source operating system.

2.2.2. Models of online collaboration

While a detailed examination of online collaboration models is beyond the scope of this study, this sub-section provides an overview of relevant literature regarding different models observed in online collaborative initiatives in order to contextualize online collaboration and its various products, including the case in hand, i.e. Ubuntu.

Haythornthwaite distinguishes between "crowds" and "communities" when it comes to the actors involved in a collaborative project, or "peer production process" in her terms (Haythornthwaite 2009: 1). Her analysis views peer production models on a continuum, with lightweight collaboration among crowds on one end, and heavyweight collaboration among communities on the other. She states that the lightweight model is "based on individual, discrete, pooled contributions" and the heavyweight model is "characterized by long-term commitment to common cause and community functions" (ibid. 9). Haythornthwaite notes that actual examples of online collaboration tend to display characteristics of both models, operating at different levels concerning different components of the collaborative project. She uses the example of Wikipedia

as it "demonstrates both light and heavyweight behaviors: lightweight from the crowds who enter, edit, and update entries; heavyweight from the inner circle of editors" (ibid. 4). This observation applies to Ubuntu as well, since it is sustained by its community of users who have a shared interest in open-source software, thus exhibiting heavyweight characteristics. On the other hand, there is no restriction on lightweight contributions from individuals with a relatively shorter-term motivation or interest, as in a single posting in the community forums to resolve a particular issue encountered by a community member.

Daoud et al. (2010) break down the essential components of user-generated content production as "Contribution Factors", which consist of a volunteer, a contribution unit, a motivation, and a collaborative environment. The authors present an initiative to create a collaborative multilingual lexical resource with an emphasis on "preliminary unvalidated data (preterminology rather than terminology)" in order to encourage volunteer participation through varied, active and passive means. The idea of collecting user input on terminology (or "preterminology") follows a similar motivation to this study in terms of acknowledging the importance of the users and the mechanisms that can facilitate the reflection of their preferences on the formation of terminology.

Folaron (2010) provides an account of various applications and uses of collaborative technologies in translation from a networking perspective, explaining the various modes of collaboration ranging from translation forums to crowdsourcing platforms in connection with the emergence of online volunteer and collaborative networks contributing to a variety of translation activities. She points out that examples of these networks operating around open-source software assume a particularly significant role by addressing the needs of linguistic communities that are not prioritized by commercial localization networks (ibid. 233).

Kelly et al. (2011) identify two bases to categorize collaborative work models: motivation and platform. On the basis of motivation, the authors identify three models. Cause-driven models are those that are mostly used for non-profit purposes, for example by non-governmental organizations. Product-driven models refer to initiatives that focus on the undertaking of tasks relating to a specific product, for example the case of Adobe recruiting contributors for localization and often remunerating them "through free products, services, or promotional merchandise" (ibid. 89). Finally, outsourcing-driven models are essentially driven by a motivation to reduce costs without prioritizing quality, as is the case with Amazon's Mechanical Turk[26], which provides a platform for seekers of online labour to meet contributors. The case of Ubuntu can be considered a mix of cause-driven and product-driven, as it represents an investment in the open source movement and constitutes a non-profit phenomenon, while involving tasks that relate to a specific software product.

On the basis of the platform used in collaborative translation, Kelly et al. (ibid. 90) identify wiki/forum-based, database-driven and full-blown collaborative translation platforms. The extent of the platform used in collaborative translation also indicates the scale of collaboration, since wiki/forum-type platforms are essentially limited to providing a means of communication, databases can facilitate further collaboration through the sharing of various types of content (e.g. glossaries, terminology databases, translation memory databases and so on), and full-blown collaborative translation platforms can integrate a multitude of functions ranging from simultaneous online collaboration to various language technology functions (terminology management, translation memories, etc.). Collaboration among members of the Ubuntu community takes place in all of the above mentioned platforms in various contexts, as in

[26] https://www.mturk.com, last accessed 24 June 2016.

feedback and discussion regarding translation in community forums, terminology databases that are shared among its volunteer translators, and its own collaborative translation platform that helps pool translation resources[27].

In their analysis of various practices in collaborative translation, Désilets and van der Meer (2011) identify five common issues: defining clear and realistic business goals, motivation, quality control, addressing the role of professionals, and parallelism and de-contextualization. Their study provides a detailed account of efforts to create a shared database of best practices in collaborative translation, and presents a set of six best practices that align with the common issues identified. First, to define clear and realistic business goals before the start of the collaborative initiative, the authors underline the importance of aligning the expectations of all stakeholders and having backup plans and check points in place (ibid. 38). Second, to address motivation, they propose a variety of mechanisms that can be used to motivate contributors, ranging from point systems to official certificates (ibid.). Third, the authors outline various methods of quality assurance that can be used in a collaborative translation setting, such as peer reviewing and voting (ibid. 39). Fourth, they suggest facilitating a contributor career path in order to provide the community with the opportunity "to grow into different roles and participate meaningfully and to the best of their ability and availability" (ibid.). Fifth, in order to address the issue of parallelism and de-contextualisation, which refers to problems arising from the fact that work is often divided into small chunks for collaborators to share, potentially leading to problems of coherence and consistency, Désilets and van der Meer propose right-sizing, i.e. finding the optimal number of collaborators depending on the size of the project in hand, as a means of ensuring appropriate distribution of workload and facilitating team work and

[27] http://community.ubuntu.com/contribute/translations/, last accessed 24 June 2016.

collaboration (ibid.). Finally, the authors point to the role of the tools and processes in collaborative contexts and show how new practices require new and improved tools and processes that are able to provide the necessary facilities and functionalities (ibid. 40). To sum up, the authors highlight critical checkpoints that can be used as benchmarks in examining online collaborative initiatives.

The framework proposed by Désilets and van der Meer can be applied to the collaborative localization of Ubuntu in order to reveal details about each issue identified by the authors and the best practices identified in response to those issues. A thorough analysis of a large-scale collaborative initiative such as Ubuntu would have to involve the study of its numerous aspects such as programming, software security and global user community networks, which are for the most part inseparable from each other and from localization. This type of analysis is outside the scope of this study. However, a brief overview of some aspects of collaborative localization is provided here in order to contextualize the localization of Ubuntu within the framework proposed by Désilets and van der Meer (2011).

In terms of issues revolving around business goals and corresponding best practices for planning and scoping as well as right-sizing, the collaborative localization of Ubuntu benefits from the open source model. This allows for unrestricted and unlimited contributions from any contributor, but also has the downside of creating coordination issues in the case of large groups (Shirky 2008). This means that anyone who has anything to contribute is free to do so, without any restrictions on the minimum or maximum quantity or quality of contribution. In terms of planning, collaboration in the Ubuntu context takes place rather freely, without any strict planning imposed on contributors, who are free to contributing on any aspect of collaboration requiring input. This may result in some inconsistencies regarding the functionality or

completeness of each aspect of the software (e.g. certain parts of the software interface not being localized or certain parts of the software not being fully functional). However, through coordination within the community, these issues are addressed relatively quickly and efficiently, as the advantage of having a vast base of contributors in the open source setting has been acknowledged as one of the major strengths of open source compared with proprietary systems that limit themselves to the contributions of their employees (Souphavanh and Karoonboonyanan 2005:3; Shirky 2008; Shirky 2012).

As for ensuring the quality of contributions, every aspect of collaboration within Ubuntu, whether it is localization, software development, updates, or software security, benefits from groups of users who are actively involved in testing, verifying and ratifying contributions suggested by others. While useful contributions are relatively quickly ratified and implemented in new versions of the software or its components, low-quality contributions are filtered out and most frequently do not make it into any new release. In the event that they do get released and cause issues, they are pointed out and addressed within the community through user feedback that is shared on the discussion forums. This constitutes a good example of how the massive number of contributions most commonly make up for the shortcomings of low-quality contributions (Surowiecki 2004:10; Howe 2006; Shirky 2008; Needleman 2011).

Contributors in open-source initiatives are most commonly motivated by a shared interest in improving and promoting open source, as well as a passion for computer programming (cf. Blum 2012: 125, 126). In addition, there are specific motivators within groups of contributors working on a certain aspect of the collaborative initiative, ranging from the acknowledgment of valuable contributions to a community-wide reputation based on a combination of the quantity and quality of contributions. As for tools and processes that are used for various aspects of

collaboration, Ubuntu relies on a wide variety of platforms that facilitate communication and coordination among contributors, ranging from software tools to communication platforms such as discussion forums.

2.2.3. Implications for translation research and practice

The most common concern shared by translators and translation scholars in light of the increasing use of various online communities for translation services seems to be quality, judging by the discussion in relevant literature (Baer 2010; Munro 2010; Désilets and van der Meer 2011; Needleman 2011; O'Hagan 2011). On the other hand, as indicated previously, most researchers seem to agree that in many cases the broader base of contributors makes up for low-quality contributions, and the expertise of professionals and the mass production capacity of the crowd are complementary to each other, particularly in situations where the workload is shared by the two groups. Jimenez-Crespo (2011) argues that quality control by users of localized products generally leads to high quality products thanks to users' awareness of expectations from the product, in addition to their knowledge of genre conventions.

Open collaborative localization offers immense potential to meet the constantly and rapidly increasing demand for global access to information and new technologies, especially for linguistic communities that may not constitute a profitable market despite their motivation to keep up with the rest of world. In the meantime, the expertise of professional translators will always be needed for "high-end services", in the words of Needleman (2011), who compares the situation to the production of popular and luxury cars: "Porsche still sells $100,000 911s while Ford cranks out Fiestas." He argues that "shunting less-interesting jobs to the assembly line workers will leave craftsmen with more time to work on the hard, interesting, and genuinely difficult tasks; and leave publishers and advertisers with enough money to pay them to do so."

While translation, both as a professional activity and as a field of research, is undoubtedly going through an important period of transformation influenced by developments in technology, the professional translation community could seize the opportunity to claim its role as the centre of expertise in interlingual information transfer. Baer (2010) writes:

> With a defined framework for appropriate and effective translation crowdsourcing, there is an opportunity to guide what has been considered a threat to the translation industry into a form that is more acceptable, and perhaps even into a positive model – seeding collaboration between amateur and paid professional translators, providing a training ground for new translation graduates, while, at the same time, expanding the material that gets translated, broadening access to information, and exposing more people to the translation process in all its complexity.

In the context of destabilizing institutions due to the shift of power from those who hold the means of mass publication to almost anyone with an Internet connection, Shirky (2008) suggests that we are currently going through the next big revolution, just like the invention of the printing press destabilizing the dominance of the Church. This time, the Web 2.0 phenomenon is destabilizing institutions that are founded on hierarchical structures that traditionally controlled information. The liberation of information is opening the doors of free and open communication and collaboration for individuals, and creating new models of providing and sharing content and services. To be able to adapt to this new setting, Shirky says, "We might as well get good at it!" (ibid.)

The road to understanding this new phenomenon creates many avenues of research in translation studies. Jimenez-Crespo (2011: 147) suggests that if collaboratively translated content

was examined in a corpus-based study, the results could help us gain a better understanding of user expectations within the framework of digital genre conventions, providing a different perspective on what is already expressed in collaborative platforms that allow users to vote on each other's translations. This suggestion aligns closely with the primary aim of this project: to analyze a corpus of discussions that relate to collaboratively localized content in order to gain a better understanding of terminological features that are present in users' communications as indications of their terminological preferences. The difference here is the focus on users' communications instead of translated or localized material. In fact, in collaborative initiatives where the evaluation of translations is based on users' votes on each other's translations, the voting process may not actually represent the linguistic choices made in authentic communications. By examining users' communications, this study focuses on actual use rather than preference indicated through voting, and it thus presents an assessment of the outcome of collaborative localization as it pertains to target language terminology.

Another question that relates to user preferences in the context of open collaboration, regardless of how they are indicated or analyzed, is the question of what exactly to do with them, and to what extent they are or should be a determinant of how to carry out the activity. Researchers who examine quality assurance in open collaboration have started exploring the issue; however, this is an ongoing debate with no straightforward conclusion or clear answers. If a certain term is preferred by the users of localized software, does that mean it is the correct term to use? It may not be. However, there is a common trait shared by open collaboration and language, and it has to do with the fact that they are both by nature interactive and constantly evolving. Thus, it would be fair to say that if a preferred term is indeed not the correct term, the collaborative dynamics are likely to identify the problem and adopt a different term if needed. It

is important to remember that even if there is an observable pattern in a collaboratively made decision, which is one of the points that this study aims to explore, there is no guarantee that the pattern will continue in future contexts. For example, term formation patterns exhibited by terms with higher rates of implantation may not necessarily maintain these rates of implantation in the future, as language evolves and terminological preferences evolve with it. The good thing is that contributors involved in collaborative initiatives are capable of running their own checks and balances to fix the problems observed in their output, since that is how collaborative initiatives function at a very basic level. This is also a reminder of one of the limitations of this project as explained in sub-section 1.3 regarding the challenges of conducting research on a subject matter that is constantly and rapidly evolving. However, shedding light on any part of the puzzle certainly plays a role in building knowledge towards understanding the phenomenon.

There is another angle to the question raised in the previous paragraph. How will complete freedom over terminological choices affect linguistic integrity? What if user preferences are at odds with the rules of the language? How does that help linguistic identity? The current era is witness to the blurring of various types of identities and borders (national, ethnic, racial and so on) with increased global mobility and interaction. These issues are entirely outside the scope of this study, but the position taken within this project sees language as an evolving social phenomenon and accepts the temporary and permanent changes in language as they are, without adhering to any inherently correct form of language (cf. Laforest 2002: 81).

2.2.4. Internet research ethics

Research on user-generated content has generated new contexts and considerations concerning research ethics, as ethical principles regarding respect for autonomy, avoiding harm and the protection of privacy require a new degree of attention and consideration in online settings. In

order to address emerging contexts for research ethics and to provide ethical guidelines for Internet research, the Association of Internet Researchers (AoIR) published a report titled "Ethical decision-making and Internet research" (Ess et al. 2002). The updated second version of the AoIR report published in 2012 underlines some key considerations regarding ethics in online research related to the involvement of human subjects, privacy and the protection of data (Markham and Buchanan 2012). With these underlying considerations, Zimmer (2010), for instance, calls into question the ethics of a 2008 study based on information collected from the personal Facebook pages of a group of Harvard University students. The author points out that a research assistant who was part of the university network was employed to collect information, despite the position taken in the study arguing that the data was publicly available (ibid. 321). This results in a problem that stems from privacy settings enabling users in the same network to access a certain amount of personal information that is in fact not available to the public at large. Therefore, the author concludes that the researchers conducting the 2008 study fail to adhere to ethical standards, and underlines the importance of the awareness of emerging ethical challenges involved in using online user data for research purposes (ibid. 323).

Warrell and Jacobsen (2014) point out the policy gap in Canada regarding Internet research ethics. Their article provides a framework for the assessment of key ethical issues that are relevant in online research settings. This framework is used here to justify the ethical position taken in this thesis. First, the researcher is not involved and does not interact with any individual user or group of users that participate in the online discussion forums which are the source of the data collected in this study. Therefore, this thesis constitutes "non-intrusive web-based research" as it is based on "the collection and analysis of existing data on the web" (ibid. 25). Accordingly, it is outside the realm of human subject research requiring an ethics review. Second, neither the

corpus nor any associated data is disseminated with this research. The corpus is saved on a password-protected personal computer, for access only by the researcher. The only excerpt from the corpus that is disseminated as part of this dissertation is presented in a screenshot on page 67, and it does not include any identifiable user information. Consequently, the anonymity and the privacy of users who participate in the discussion forums that are the source of the data collected in this study are fully protected.

2.3. Terminology

This section first summarizes the evolution of terminology theory in sub-section 2.3.1 and presents a theoretical overview of term formation in sub-section 2.3.2. Terminometric analysis, which forms the foundation of the methodology followed in this study, is discussed in sub-section 2.3.3.

2.3.1. Theories of terminology

The traditional theory of terminology spearheaded by Eugen Wüster (the Vienna school) advocates for terminology standardization through a concept-based, onomasiological approach, going from concept to term, based on the ideal that terminology should provide a perfectly efficient tool in communicating specialized concepts without ambiguity. This approach starts by isolating a concept and its meaning, and finishes by assigning it a linguistic designation. The aim is to eliminate ambiguity in order to achieve monosemy and univocity, in line with one of the central arguments of traditional terminology theory, i.e. the view that the concept is universal and the difference is in the language used to express the concept (Cabré Castellvi 2003: 165-167). The principle of one-to-one correspondence between concepts and terms leaves little or no room for the analysis of context. Indeed, if one-to-one correspondence between concepts and terms is

achieved, then contexts in which terms are used have little relevance except for providing examples of usage. This, among other reasons, is why the Vienna school has received a considerable amount of criticism from a variety of fields especially starting in the 1990s. Ultimately, this led terminologists and researchers of terminology to the realization that "the traditional theory was unsatisfactory for the description of the real data in all their complexity" (ibid. 187).

Although the traditional theory still constitutes the foundations of several large-scale terminology initiatives including the two largest terminology databases in Canada (TERMIUM[28] and the GDT[29]), the need to bring terminology theory up to date with current needs and practices of terminologists motivated the development of new theories that build on some of the foundations of the Vienna school while using a semasiological approach, moving from terms to concepts. The basis of this new perspective has been the observation of terms in context, which has been facilitated by the increasing use of computers to analyze large amounts of textual data in the form of corpora that serve as evidence of language use and that can be used to test hypotheses about language. An increased interest in corpus-based terminological studies and the analysis of terms in context has shifted the focus from the normative nature of traditional terminology theory toward new theories with a descriptive stance.

One of these newer perspectives on terminology theory is the textual approach discussed by Bourigault and Slodzian (1999). Their theory builds on the increasing variety of situations in which terminology work is done, and the increasing variety of terminological data used by computer programs (ibid. 29). With so many new and different applications, the need arises to

[28]TERMIUM is the Government of Canada's official online terminology database, http://www.btb.termiumplus.gc.ca/ (last accessed 24 June 2016).
[29]Le grand dictionnaire terminologique (GDT) is the official online terminology database maintained by the Office québecois de la langue française (OQLF), http://www.gdt.oqlf.gouv.qc.ca (last accessed 24 June 2016).

adapt terminology research and management methods to the requirements of these applications. The textual approach to terminology is based on the idea that there may be as many ways of representing the terminology of a specialized domain as the number of situations in which the terminology will be used (ibid. 30). This introduces a significant amount of flexibility into terminology research and management compared to the traditional approach, which imposed strict normative guidelines. It also illustrates the evolution of the field from prescriptive methodologies towards descriptive ones. The corpus as the primary linguistic evidence of the exchange of knowledge that takes place within a community has become the starting point for most terminology work, thanks to the capability of corpus analysis tools to process large amounts of data quickly (ibid. 30). The textual approach therefore is interested in corpus-based analysis of terminology and calls for methodological and theoretical changes that can pave the way to terminology research suited to its applications in specific situations.

Cabré Castellvi's communicative theory is another example of a theory of terminology that gives priority to the specific situations in which terminological units are used (2003: 186). As a reaction to the restrictions of the traditional theory that prove problematic in the face of real-life linguistic complexities such as polysemy and synonymy, Cabré Castellvi's argument focuses on the "multidimensional nature of terminological units" – "the cognitive (the concept), the linguistic (the term) and the communicative (the situation)." (ibid. 187). The communicative theory of terminology acknowledges the impossibility of reaching an absolute and complete account of terminology due to its integral proximity to the complex phenomenon of language. This theory offers an explanation of terminology that aims to construct access points for the examination of terminology rather than claiming to achieve a full account of all terminological phenomena within a single theory.

Another terminology theory that takes into account the importance of context and perspective is the socio-cognitive theory of terminology, as discussed by Temmerman (2000). Using corpus analysis as its starting point, this approach builds on the importance of terms in context, arguing that in authentic situations where terminology is used, concepts tend not to have well-defined boundaries as suggested by traditional terminology theory (ibid. 59). Due to the instability of the concept in authentic situations, the socio-cognitive theory is based on units of understanding ("unités de compréhension") and terms in their textual environment, which does away with the notion of universality of concepts (ibid. 60). Temmerman also acknowledges that the definition of a term can vary depending on the context in which it is used, and therefore polysemy and synonymy are inevitable linguistic phenomena (ibid. 61) that the field of terminology cannot escape.

The socio-cognitive theory of terminology is one of the building blocks of modern terminology research and practice. In contrast to traditional terminology theory, it challenges the clear separation of the term and the concept and lays the foundations of understanding terminology not only in its immediate textual context but also in a wider context that includes cognitive and conceptual relations between the term and the concept. Temmerman discusses the influence of cognitive models on terminology and naming mechanisms in particular. She suggests that the socio-cognitive approach can lead to a better understanding of neologisms and the evolution of terms and categories through an analysis of terminology in relation to the surrounding socio-cognitive context, which motivates the formation and evolution of terms (ibid. 61, 62). This study draws in part on the socio-cognitive approach as it analyzes terminology in the context of discussions that take place within an open-source software community and examines the relationship between term formation and implantation. However, the examination

of cognitive models that operate within this context is outside the scope of this project, and the results of the study reflect a snapshot of term implantation rather than the evolution of terms or term implantation over time.

The relatively more recent socioterminological approach introduces yet another perspective on the role of context in terminology, with a focus on social context and its role in the formation, dissemination and use of terminology. Gaudin (2007: 27) argues that the classical approach that relied on fixed terminological categories and monolithic terminological structures proves problematic in practice, since it relies on the assumption that concepts are universal and ignores the socio-historical context. He presents the beginnings of socioterminology based on the example of the status of French language in Québec. Following the Quiet Revolution of the 1960s, the efforts to establish French as the language of business in the province led to the realization that prescriptive attitudes do not produce sustainable outcomes and that the observation of terminology in use is key for the understanding of terminological processes within a linguistic community (ibid. 28). According to Gaudin, this realization, which at the time mostly concerned sociolinguistics, plays an important role for a linguistic community to establish its linguistic independence, and later on it gave rise to the field of socioterminology. Terminometrics, which is discussed in more detail in sub-section 2.3.3, is a field that has emerged as a measurement tool to assess the outcome of language planning in the context of socioterminological research looking into the dissemination and implantation of terminology.

This study relies for the most part on the socioterminological approach. It examines terminology as it is used within the social context of online collaboration, and searches for answers regarding the influence of this particular context on the dissemination and use of terminology within a particular domain. It sets aside prescriptive views, accepts language and

therefore terminology as phenomena that are constantly evolving, and attempts to reach a better understanding of terminological choices in the context of open collaborative localization. Finally, the methodology used in the study is built upon terminometric research methods, which constitute an important component of socioterminological research.

2.3.2. Term formation

Unlike the formation of words in language for general purposes, which appears to take place relatively arbitrarily, the formation of terms tends to exhibit more regularities or patterns (Sager 1997: 25). This is not a coincidence, since terminology as science strives to be systematic in the naming of concepts to achieve transparency and consistency. One of the interests of terminology research is to examine how linguistic communities formulate terminology to express concepts, and to assess such regularities, patterns and other factors that play a role in term formation. This is a means of understanding the rules that direct or motivate naming conventions, and of formulating guidelines for future designations (Sager 1990: 61). One of the reasons why terminology is more likely to follow rules and patterns is that it is theoretically the result of the work of subject-field experts, terminologists and/or linguists (Maurais 1993: 112). However, as mentioned in the previous sub-section (2.3.1), recent terminology research has shown that the creation and evolution of terminology can be much more difficult to predict or prescribe in practice (Bowker and Hawkins 2006: 83). In the face of the sometimes unpredictable nature of language and how terminology operates within it, Pavel suggests:

> Finding out the causes, the patterns or regularities hidden behind such apparent randomness is one of the new tasks facing terminologists. Its neglect hinders standardization efforts, leads to inconsistencies among vocabularies dealing with the same field of expertise, and partially

explains their incompleteness. (Pavel 1993: 24)

Many researchers have observed that the formation and dissemination of terminology is in constant interaction with a variety of factors related to human nature, community, history, experience and knowledge, culture, memory, perception of reality, reconceptualization and so on (Diki-Kidiri 2007: 14, 15). Recent research on terminology, as documented in Marcel Diki-Kidiri's work on cultural terminology (ibid.), reveals how complex a role terminology plays within language. For the purposes of this study, while acknowledging the complexity of term formation and the transfer of concepts between two linguistic communities, the terminology component of the theoretical framework is mainly built upon Sager's (1990) observations on term formation.

Sager (1997: 27) observes two essential instances of term formation: primary and secondary term formation. He describes primary term formation as terminology creation that is more or less simultaneous with concept formation (ibid.). In other words, this refers to a situation where a concept originates within a linguistic community, and is consequently named by the same community. In most cases, such term formation is carried out by means of linguistic instruments at the disposal of the concerned language, without recourse to external resources, and is therefore generally a monolingual process. The main focus of this study, however, is on secondary term formation. Secondary term formation occurs when the designation for an existing concept needs to be renewed within a single language, or when an existing concept is transferred from one linguistic community to another and its designation therefore needs to be accommodated by the target language (ibid.). Secondary term formation resulting from the importation of concepts from different linguistic communities forms the focal point of this study, as it is a common occurrence in the contexts of technology and localization. As noted by

Rousseau (1993: 40), "the extremely technical nature of today's world creates a considerable, ever-increasing need for terms. These technologies are, for the most part, imported, and their terminology is in English." Although this observation is now more than two decades old, it retains its relevance in this age of rapid and constant technological innovation.

Sager, who notes that the importation of concepts from other linguistic communities is a natural component of the evolution of language, presents several examples of how the influx of foreign terminology can be balanced by indigenous production of texts and materials, as well as teaching and training in the indigenous language (Sager 1990: 82-4). Such linguistic activity serves as a means of incorporating the knowledge within the linguistic community, allowing it to apply its own linguistic mechanisms. However, in certain specialized fields where there is an imbalance between languages that are dominant (e.g. English as the main language of scientific and technological innovation) and those that are somewhat dependent on external sources of information in such fields, the lack of conscious effort for such linguistic activity can lead to various problems concerning language and terminology. Numerous examples can be observed in all parts of the world, from Bouchard's (1989) account of the centuries-old efforts of Francophones in Québec to preserve their linguistic identity, to a study carried out in Turkey on the effects of foreign language education on native language skills (Aydın et al. 2004). Despite a wide variety of differences between the two contexts such as geographical distance and socio-historical, cultural and linguistic factors, they both involve linguistic communities under the influence of English. Both studies present evidence of linguistic influence and subsequent resistance, in varying forms and levels, and demonstrate the outcome of such situations. The former (Bouchard 1989: 87) refers to calls for a conscious effort by individuals and the community to protect the French language from Anglicisation. The latter (Aydın et al. 2004)

demonstrates the adverse effects of university education carried out in a foreign language (English) on native language (Turkish) skills. Deteriorating native language skills undoubtedly influence terminology development mechanisms, which are likely, then, to reflect the influence of dominant foreign languages – for example in the form of Anglicization. Such research indicates the significance of interlinguistic influence and its possible impact on term formation.

The transfer of technological knowledge through terminology presents various challenges for the target language. One of these challenges is standardization of terminology, which can be hindered due to different translations and unstable terminological frameworks. Sager elaborates on the issue of standardization in the context of the terminology of technology as follows:

> The terminology of technology […] is volatile both in its form and existence because of changes in materials, methods of production, design, etc. This lack of stability is accentuated in transferred terminology, i.e. terminology created by secondary term formation from concepts borrowed from another linguistic community. The existence of several methods of secondary interlingual term formation, e.g. direct borrowing, loan translation, paraphrase, parallel formation/constituent recreation, adaptation, complete new creation, which may be used simultaneously or sequentially, provides the occasion for several alternative or competing new terms and it may therefore be several years before a terminology stabilizes. (Sager 1990: 82)

The main motivation behind the selection of Sager's framework for term formation patterns for the purposes of this project is that it provides a clear set of criteria that facilitates the identification of characteristics pertaining to each pattern, thus facilitating a discussion of

potential correlations between these patterns and term implantation rates (see section 3.1 for the explanation of term formation patterns and how they are used in the methodology applied in this study).

The next sub-section presents a summary of terminometric research methods that form the basis of the methodology used in this project in measuring implantation rates for terms exhibiting each formation pattern.

2.3.3. Terminometrics

Terminometric research emerged out of the need to measure term implantation in order to evaluate the results of language planning efforts and to be able to improve terminological suggestions and guidelines according to factors seemingly affecting term implantation. This sub-section presents a brief overview of language planning before discussing terminometrics and methodologies used in terminometric studies.

Humbley (1997: 261) observes that language planning is commonly employed as "a solution to problems engendered by multilingualism and/or language standardization linked with economic development." It usually takes the form of terminological recommendations, sometimes legally imposed with varying levels of ramifications by government bodies, and also implemented by ministries of education, influential official institutions, large companies and dictionary publishers (ibid. 262). Thus, language planning is aimed at providing an official point of reference for terminological choices, and terminometric research serves to evaluate these choices and determine how effective language planning efforts are in shaping or influencing them.

Terminometric studies have mainly been used in the context of assessing language planning initiatives and applied to institutional communications (Quirion 2003a: 31) based on the

understanding that only institutions could determine the linguistic situation within a linguistic community over the long term (Corbeil 1980: 116). This may have been an indisputable observation for as long as institutions had control over the production and dissemination of substantial amounts of linguistic content through internal and external communications. However, widespread Internet use and the digitization of information have changed the way content is produced and distributed. The increased potential influence of individuals or groups on language, facilitated by the technical possibilities offered by the Internet that eliminate the need for major publishing and distribution mechanisms, creates a need to question the previously accepted role of the institution in existing literature as the proof and determinant of language use in a linguistic community.

A large part of the existing research that has evaluated the implantation of terms suggested as part of a language planning initiative are based on surveys or questionnaires about language use among subject-field professionals. For example, Daoust (1995) examines factors influencing terminological choices through interviews, surveys and field observations. She acknowledges that claimed terminology use does not reflect actual use, yet asserts that it represents social consensus among the community (ibid. 280). The fundamental drawback of this approach is that surveys and questionnaires are likely to elicit non-spontaneous and potentially biased responses. Quirion (2003a: 31) emphasizes the importance of distinguishing between usage and awareness. In essence, there is a need to differentiate between active and passive knowledge of terminology, and between actual and claimed usage of terminology. This is important given that claiming use of certain terms in a survey or questionnaire is no proof of actual terminology use in authentic contexts. These points underline the need for objective and reproducible data with a minimal involvement of arbitrary factors in order to ensure

comparability among different studies, which is essential for diachronic studies in particular (ibid. 34). For these reasons, Quirion (ibid.) suggests a terminometric survey protocol to calculate the relative frequency of terms used to express a certain concept as a measure of implantation ("implantation coefficient"[30]) based on a corpus that is representative of authentic language use. The protocol consists of six steps: 1) determination of the subject area, 2) selection of the terms to be measured, 3) construction of the corpus, 4) determination of sample size, 5) data collection and processing and 6) the presentation and discussion of results (ibid. 35-42). Some of the studies that are based the application of this protocol are outlined below.

Quirion (2003b) uses this protocol to assess the status of terminology in the transportation domain in Québec. Quirion and Lanthier (2006) use the protocol in order to verify principles of term selection suggested, but previously not factually tested by terminology theory. Quirion (2004: 194) outlines terminological, socioterminological and procedural factors that may influence term implantation by studying both explicit and implicit mentions of such factors in terminology literature. Terminological factors refer to those intrinsic to the term, such as concision, conformity with the rules of the language and derivability (ibid. 195). Socioterminological factors are those shaped by social elements affecting terminology and language, such as the role of the speaker, linguistic insecurity and linguistic policy (ibid. 196, 197). Finally, procedural factors relate to the methods and processes that apply to the formation of terminology, such as the method of terminology compilation or the timing of the release of terminology (ibid. 197). Quirion's observations suggest that the participation of the potential users of terminology in the decision-making process would encourage the adoption of

[30] Based on Quirion's (2003a) formula, the *implantation coefficient* (IC), which is represented by a value from 0 to 1, represents the relative frequency of a term used to designate a concept in relation to all other competing terms that designate the same concept.

terminology by the users of language in a specialized domain (ibid.). Quirion and Lanthier (2006) conduct a study where the authors develop a methodology based on terminometric analysis to assess the influence of a selection of factors frequently cited in terminology literature. In particular, the authors focus on the influence of terminological factors that "concern the intrinsic characteristics of the terms" by analyzing "terms at opposite ends of the implantation spectrum" (ibid. 111).

More recently, the same terminometric protocol is applied in Karabacak's (2009) study of terminology suggested by Turkey's official language institute, the Turkish Language Society. With a focus on terms in the economic domain, the author analyzes a corpus of newspaper articles and develops suggestions for more efficient terminological planning. The study combines corpus analysis and a survey on the attitudes of journalists writing for Turkish newspapers towards language planning, in an attempt to better understand the factors motivating their choice of terminology and identify ways to bridge an apparent gap between officially sanctioned terminology and terminology that is used in the field. This combination constitutes an original approach that leads to observations based on both quantitative data from the corpus and the unique insight provided by subject-field experts. The author concludes that the lack of standardization and the lack of a credible authority, despite the existence of the Turkish Language Society, seem to be the two main obstacles facing terminology planning in Turkey. Based on the survey, he suggests more expert involvement as the way forward for improved efficiency in language planning (ibid. 168).

The studies summarized above use Quirion's (2003a) protocol without any substantial modification, and they are carried out in language planning contexts as intended by the original protocol. This thesis, however, adapts this protocol for use in a different context, where the

emphasis is put on language use by the community as opposed to the institution. The next chapter explains this adaptation in detail and lays out the manner in which the protocol is applied in the methodology used in this study (see section 3.2).

3. METHODOLOGY

This chapter presents the methodology followed in this thesis project. It should be noted that an initial and basic version of the methodology was applied in a small-scale pilot study on the terminology of the Turkish version of the open-source office suite OpenOffice (Bilgen 2012).

In the pilot study, the discussion forums of the local Ubuntu community in Turkey were manually scanned for titles specifically related to OpenOffice. The inclusion of these titles resulted in a corpus of about 20,000 words, with 75 topic titles and approximately 600 postings at the time of collection. Concepts designated by competing terms[31] within the corpus were recorded, in addition to terms designating the same concepts found in major Turkish dictionaries of information technology[32], and in the Turkish version of Microsoft Office as the mainstream office suite[33]. Implantation coefficients (IC) were calculated for each term in line with the methodology suggested by Quirion (2003a), followed by an etymological examination of each term to identify how they were formed[34]. The study revealed choices of terminology in discussion forums that sometimes differed from MS Office terminology, Open Office terminology, and/or terminology stored in dictionaries. Table 1 below summarizes the data collected in the pilot study.

---

[31] *Competing term* refers to one of the terms that is used to designate a notion which is not designated by a single term. This applies to most terms in the context of this study due to the predominance of non-standardized terminology in the subject field in question (cf. Sager 1990: 82, Lepouras and Weir 1999). The term *competing* is deliberately used instead of *synonymous* since it allows for the inclusion of terms in different languages (English and French terms in this study), and carries the implication of non-standardization.

[32] The two following specialized dictionaries were consulted for this purpose:
SANKUR, Bülent (ed.) (2008) *Bilişim Sözlüğü [IT Dictionary],* Istanbul: Pusula.
TBD Bilişim Terimleri Karşılıklar Sözlüğü [IT Terms and Equivalents Dictionary], Informatics Association of Turkey, http://www.tbd.org.tr/index.php?sayfa=sozluk, last accessed 24 June 2016.

[33] MS Language Portal was used to check for Turkish terms used in Microsoft products to designate each concept. https://www.microsoft.com/Language/, last accessed 24 June 2016.

[34] The following etymological dictionary was consulted for this purpose to assist in the understanding of term formation processes for each term:
NİŞANYAN, Sevan (2009) *Sözlerin Soyağacı. Çağdaş Türkçe'nin Etimolojik Sözlüğü. [Genealogy of Words. An etymological dictionary of modern Turkish.]* Istanbul: Alfa. Online version: www.nisanyansozluk.com, last accessed 24 June 2016.

Concept	Term	IC	Source(s)
default	varsayılan	0.24	Open Office, MS Office, Corpus
default	öntanımlı	0.76	Corpus
font	yazı tipi	0.31	Open Office, MS Office, Corpus
font	yazıyüzü	0.31	Corpus
font	font	0.38	Corpus, IT dictionaries
format	biçim	0.58	Open Office, MS Office, Corpus, IT dictionaries
format	format	0.42	Corpus, IT dictionaries
page setup	sayfa biçimi	0.00	Open Office
page setup	sayfa yapısı	1.00	Corpus
page setup	sayfa düzeni	0.00	IT dictionaries
spell check	imla denetimi	0.00	Open Office
spell check	imla kontrolü	0.50	Corpus
spell check	spell check	0.50	Corpus
spell check	yazım denetleme	0.00	MS Office
spell check	yazım denetimi	0.00	IT dictionaries
spreadsheet	hesap tablosu	0.35	Open Office, Corpus
spreadsheet	hesap çizelgesi	0.24	IT dictionaries, Corpus
spreadsheet	spreadsheet	0.24	Corpus
spreadsheet	çalışma sayfası	0.18	Corpus
spreadsheet	elektronik çizelge	0.00	IT dictionaries
tab	sekme	0.70	Open Office, MS Office, Corpus
tab	tab	0.30	Corpus
version	sürüm	0.93	Open Office, MS Office, Corpus, IT dictionaries
version	versiyon	0.07	Corpus

Table 1 - Data collected in pilot study (Bilgen 2012)

For example, for the concept *font*, three terms were used in the corpus with similar implantation rates (*yazı tipi,* IC=0.31; *yazıyüzü,* IC=0.31; *font,* IC=0.38). Two of these terms were different from the term used in Open Office and Microsoft Office (*yazı tipi*), and one of them was found neither in the interfaces of these two office suites, nor in the IT dictionaries consulted (*yazıyüzü*). For another concept, *default,* the term with the higher implantation rate (*öntanımlı*, IC=0.76) was

different than the term used in both office suite interfaces (*varsayılan,* IC=0.24), and was not found in the IT dictionaries consulted.

Perhaps due to the low number of terms analyzed in the pilot study (24 terms designating 8 concepts), the etymological examination did not reveal any correlations between term formation and term implantation that could be generalized or applied in different contexts. However, the variation in terminological choices indicated that a more exhaustive application of the methodology on a larger scale might show whether such correlations exist.

The methodology applied in this study has been expanded based on the pilot study in consideration of the observations and feedback received from scholars and peers, and in light of the larger scope of the doctoral thesis project. Section 3.1 explains how Sager's (1990) framework of term formation patterns is used to identify patterns exhibited by terms that are selected for analysis. Section 3.2 outlines the terminometric research protocol presented by Quirion (2003a), and explains how it is adapted for the scope and purposes of this project.

3.1. Identification of term formation patterns

This section explains how Sager's (1990) framework for term formation patterns is used in this study for the identification of patterns observed in terms selected from the corpus. While Sager's framework is preserved for the most part in the methodology, the definitions of some patterns are slightly adjusted for the purposes and within the scope and limitations of this study.

The process of identification of term formation patterns was carried out three times by three individuals (including the author) in order to improve consistency and objectivity as well as accuracy by native speakers of French. The two Francophones who assisted the author (who is not a native speaker of French) in this task are professionals in translation and communication. These individuals were presented with the definitions of each term formation pattern as outlined

in the following sub-sections, and were asked to identify any pattern(s) they observed in each term. There was a minimal number of discrepancies among the three instances of identification of patterns, in which case the pattern(s) identified in two out of three instances were retained. Nonetheless, it is worth acknowledging that, despite efforts to minimize the chances of manual error and avoid subjectivity, the outcome of the identification process is still inevitably limited by the definitions of term formation patterns as presented in this document, the interpretation of these definitions by each individual carrying out the pattern identification process, and how each individual applies these interpretations in light of their own personal judgment and linguistic intuition.

The sub-sections below define the characteristics of term formation patterns, which serve as the basis for pattern identification. Examples of terms examined in the study exhibiting each pattern are provided where applicable.

3.1.1. Use of existing resources

Use of existing resources within the same language is a common phenomenon in term formation. Examples include the use of analogies with existing and/or familiar designations, which sometimes "extend the meaning of an existing term to embrace that of a new concept, e.g. ‘spaceship' or ‘aircraft'" (Sager 1990: 71). In the context of Internet use, for example, the term *browse* (*parcourir* in French) represents a specialized concept that is designated by the use of an existing lexical unit.

In order to avoid any possible ambiguity and redundancy between the use of existing resources and the modification of existing resources (explained in the following sub-section), only terms that constitute the use of existing resources in their original form are considered under use of existing resources. This includes components of compound terms that are formed through

the use of existing resources, in which case the term is identified as exhibiting multiple formation patterns. For example, the term *remote desktop* is formed through compounding (explained in sub-section 3.1.2); it includes the component *remote* which constitutes the use of existing resources, and *desktop* which is also a compound term made of *desk* and *top*. In addition, semantic neologisms, i.e. new meanings assigned to an existing form to designate a new concept, are also based on the use of existing resources and are considered as part of this pattern rather than new creations, as new creations are exclusively used to designate formal neologisms, i.e. new forms created to designate a new concept (see sub-section 3.1.3.1 below). For example, the terms *script* and *clavier,* which are both examples of fully implanted terms (see Appendix 3), are identified as terms exhibiting the use of existing resources.

3.1.2. Modification of existing resources

According to Sager (1990: 72), the modification of existing resources is the most common method in term formation. The following explains different means of modification, with examples of terms examined in this study.

- **Derivation or affixation**: the addition of affixes at the beginning (prefixation) or the end (suffixation) of lexical units (e.g. *interopérabilité*);
- **Compounding**: the combination of existing lexical units into new ones to form new simple (single-word) or complex (multi-word) terms (e.g. *environnement de bureau,* which also exhibits the use of existing resources);
- **Conversion**: the use of a lexical unit in a different grammatical category than its original (e.g. *to google*);[35]

[35] Since the focus of this study is on formation patterns, which relate to formal characteristics rather than conceptual distinctions that take precedence in terminology, terms that are used in multiple grammatical categories are grouped

• **Compression**: the use of shortened forms such as acronyms, abbreviations or clipped forms of existing lexical units (e.g. *courriel,* from *courrier électronique*).

The term *blog* (*blogue* in French) serves as an example of several means of modification mentioned above (compounding, conversion and compression). It is a compound term made of *web* and *log*. It is compressed, from *weblog* to *blog*. Its use in multiple grammatical categories is an example of conversion.

3.1.3. Creation of new lexical entities (formal neologisms)

The creation of new lexical entities, or neologisms, is common in the field of technology and is a result of "the need for the unique naming of new concepts" (Sager 1990: 79). Sager identifies two types of neologisms: new creations and borrowings from other languages (ibid.).

3.1.3.1. New creations

New creations refer to the creation of entirely new lexical entities, which may refer to brand new forms created to express new concepts (formal neologism – e.g. French term *pourriel*, which is an equivalent to *spam*), as well as new semantic value assigned to existing forms, also known as semantic neologism. As seen in the previously mentioned example of *script*, for the purposes of this study, a semantic neologism is studied under the use of existing resources (explained above in sub-section 3.1.1). Thus, the identification of new creations within the framework of this study only includes formal neologisms that appear to express new concepts in specialized fields.

---

together under an overarching concept. For example, *télécharger* and *téléchargement* are grouped together as terms relating to the overarching concept *download.* Consequently, when the corpus is queried for the character string *télécharg,* occurrences of *télécharger* (and conjugated forms) and *téléchargement(s)*, for example, are counted as terms relating to the overarching concept *download.* This is a methodological decision based on the priority accorded to examining formal characteristics of terms, and the consideration of the benefit of distinguishing between grammatical categories in each occurrence of a term against the time required to carry out this type of disambiguation, which is essential in terminology work but not necessarily so within the scope and limitations of this project. The disambiguation process followed in this study, as described in sub-section 3.2.3, involves the verification of the stem of the term and the overarching concept, regardless of grammatical category.

Many terms enter language as new creations at a certain point in time and there are several ways of determining whether a lexical item is considered *new* or not, such as assessing the presence of such items in dictionaries, their common use in daily language and their occurrences in published material (cf. Hamelin 1995; Quirion 2011). However, as previously explained at the beginning of section 3.1, the identification of all patterns including neologisms within the limits of this study rely on the definitions presented in this document and the way these definitions are interpreted and applied by each individual using their judgment and linguistic intuition during the identification of patterns.

3.1.3.2. Borrowing

Borrowing is a common method of secondary term formation, in which the designation of a concept in the source language is borrowed by other linguistic communities. Borrowings are likely to follow a "nativization process" where borrowed terms may go through varying levels of "alteration of phonological, morphological and orthographical forms in conformity with systems in the target language" (Sager 1990: 85). Sometimes, borrowed terms take the form of translations, which Sager calls "loan translations" (ibid. 86). The following explains direct and adapted borrowings, and loan translations.

- **Direct and adapted borrowings***:* Direct borrowings refer to terms that are borrowed in their original form from the source language. Adapted borrowings go through one or more of the above-mentioned alterations to conform to target language conventions on varying levels. For instance, the concept *blog* is designated in the corpus examined in this study by both a direct and an adapted borrowing. The use of the English term *blog* in French is an example of a direct borrowing, whereas the French term *blogue* is an example of adapted borrowing.
- **Loan translations***:* Loan translations constitute a relatively less visible type of borrowing, as

such terms involve "total and partial translation of phrases and compound terms" (Sager 1990: 86). In this case, the term may appear entirely nativized; however, its origin can relatively easily be tracked to the source language as it is translated, not formed through term formation mechanisms proper to the target language. (e.g. *carte-mère,* from the English term *motherboard*).

3.2. Terminometric research protocol

As previously mentioned in sub-section 2.3.3, Quirion (2003a) establishes a six-step terminometric survey protocol to measure implantation rates for competing terms, consisting of the following stages: 1) determination of the subject area, 2) selection of the terms to be measured, 3) construction of the corpus, 4) determination of sample size, 5) data collection and processing and 6) the presentation and discussion of results (ibid. 35-42). This protocol constitutes the core of the methodology followed in this thesis project. However, it is adjusted for the purposes of the project, mainly due to the fact that this study does not entirely fit the characteristics of the terminometric research contexts for which the protocol was originally developed.

Unlike conventional terminometric research, which sets out to measure a set of terms in a corpus of institutional communications, this project involves the construction of a corpus consisting of the communications of software users, and the identification of terms to be measured from within the corpus. Therefore, the order of stages 2, 3 and 4 as described by Quirion (2003a) are modified for the purposes of this study. The modifications are justified below in the respective sub-sections (3.2.2, 3.2.3 and 3.2.4), together with the explanation of each step and how it is adapted for this project.

3.2.1. Determination of subject area

Conventional terminometric studies focus on one specialized subject area in order to limit their scope and improve their efficiency. With the same goal, this study is limited to the terminology used in discussions relating to a specific software type: desktop operating systems. Operating systems are selected as the software type, since all software applications require an operating system to run on. Therefore, they are widely used by a variety of users, and it could be argued that the terminology of operating systems constitutes the basis of software terminology. In addition, the widespread use of operating systems creates extensive discussion among users, which is essential for the purposes of this study regarding the evaluation of term implantation in the discussion within the community.

According to StatCounter (2013), a website that provides statistical information on Internet users based on data obtained through page views for approximately 3 million websites, the top seven desktop operating systems in order of popularity at the early stages of this project were Windows 7, Windows XP, Windows Vista, Mac OS X, Windows 8, Linux, and iOS. Linux is the only operating system in this list that is created, distributed and maintained by volunteer users. Since the main focus of this study is user involvement in localization, the main source of terminology that is studied consists of discussions related to a Linux-based operating system. Linux is available in various distributions that operate on the same basis, but differ slightly in terms of technical properties. According to DistroWatch (n.d.), the top ten most widely used Linux distributions are Linux Mint, Ubuntu, Fedora, Debian GNU/Linux, openSUSE, Arch Linux, PCLinuxOS, CentOS, Mageia, Slackware, and FreeBSD. The website specifies that there are no official numbers to support this top-ten list or to put the operating systems on the list in a specific order (ibid.). This is due to the impossibility of tracking the real number of Linux users,

since Linux is entirely free to obtain and (re-)distribute. Due to its popularity among Linux-based operating systems, and due to the availability of community discussion forums relevant for the purposes of this study, Ubuntu was chosen to represent Linux-based open-source operating systems. The community discussion forums not only constitute a valuable source of linguistic evidence for studies such as this project, but also play an important role in developing, maintaining and improving the software by providing a platform for high-quality "user-to-user" support (Lakhani and von Hippel 2002: 923).

On the other hand, Windows being the most popularly used operating system represented by multiple versions in the above-mentioned StatCounter data, the study includes terminology from Microsoft's official terminological database, the MS Language Portal[36]. Thus, the study also assesses the presence of mainstream, commercial software terminology in the discussion of non-mainstream, non-commercial, open-source software. The processes followed in the construction of the corpus and the selection of terms to be measured are explained in the following sub-sections.

3.2.2. Construction of the corpus

As indicated in the beginning of section 3.2, the order of the steps suggested by Quirion (2003a) for the selection of terms to be measured and the construction of the corpus is reversed in this study. Unlike conventional terminometric studies that are based on assessing the implantation of a specific set of terms that are part of language planning efforts, there is no pre-determined set of terms in the case of this study. Instead, the terms are selected from a corpus that serves as evidence of the terminology used in relevant discussion.

One of the advantages of working with an open-source operating system such as Ubuntu

[36] http://www.microsoft.com/Language/, last accessed 24 June 2016.

is that its users and developers communicate on online platforms such as discussion forums and mailing lists, which present the possibility of easily accessing databases of these communications and creating corpora that can be used in their analysis. An interactive map displaying the various local Ubuntu communities and listing their websites can be accessed on the Ubuntu Local Community (LoCo) Team Portal[37]. The participation of users who actively take part in the development and localization of Ubuntu in these discussion forums constitutes an interesting case. Some of the same users posting on the discussion forums may be involved in the localization efforts, and therefore their choice of terminology may not be entirely impartial. Also, among the users who actively take part, there may be those with a higher level of influence due to various possible factors such as being involved with the community for a longer period of time or having a stronger reputation among members of the community. However, considering the size of the corpus and the number of postings from various users, any potential influence that any particular contributor or group of contributors may have on terminology use across the entire corpus is deemed to be negligible.

The corpus used in this study consists of communications of Ubuntu users in Québec on Ubuntu-Québec discussion forums[38]. The archives of these forums are available on the Ubuntu-Québec website and cover all postings since December 2005[39]. In order to limit the discussion to relatively recent issues and to limit the size of the corpus so that it is manageable on an average desktop computer using a basic corpus analysis tool, forum postings shared within the six-year period covering 2008 to 2013 (from January 2008 to December 2013 inclusively) were included in the corpus. Further details on the characteristics of the corpus are explained in the following

[37] http://loco.ubuntu.com, last accessed 24 June 2016.
[38] http://www.ubuntu-qc.org/drupal/forum, last accessed 24 June 2016.
[39] https://lists.ubuntu.com/archives/ubuntu-quebec, last accessed 24 June 2016.

sub-section, as they emerge from steps followed in the selection of terms.

3.2.3. Selection of terms to be measured

Unlike a conventional terminometric survey, this project does not set out with the goal of measuring the implantation of a specific set of terms. In order to obtain a set of terms to be measured, a sample taken from the corpus is manually examined with the goal of identifying competing terms designating pertinent concepts. This sub-section explains the process leading to the selection of terms to be measured. It should be noted that the terms examined in this study are selected from within the discussions of the Ubuntu-Québec community, and not from the terminology that is necessarily found in any integral part of Ubuntu itself, such as the user interface, help documents and so on, although a large amount of overlapping terminology exists between the terminology that is used in the discussion forums and the terminology that is present in various parts of the operating system. In light of the important role assumed by discussion forums in the context of open-source software, the stages explained below lead to a selection of terms that is considered to be representative of some of the key concepts discussed within the community that relate to Ubuntu, and reflective of the terminology that is used in and around this operating system.

First, all discussion topics within the corpus are arranged in a table indicating discussion topic title, number of messages posted per topic, and terms relevant to information technology observed in topic titles. This facilitates the observation of basic quantitative information regarding the corpus, and the assessment of topics of discussion and their popularity on the basis of topic titles and the number of messages posted under each title. Recurrent titles (i.e. titles that are posted in different times, sometimes by different users with a slightly different wording, for the discussion of the same topic) are consolidated under a single title to reflect the overall

discussion of the particular topic throughout the six-year period covered in the corpus[40]. After the consolidation of recurrent titles, the total number of topics of discussion is 1,612, and the total number of messages posted is 8,880[41].

Out of the total number of topics of discussion, the 734 topic titles that include terminology from the information technology domain are separated from those that do not include any relevant terminology. Those generic titles that are excluded from the selection are easily identifiable as they include discussions of social and/or promotional events, announcements, news, generic questions, and so on. Focusing on titles that reflect relevant terminology use can facilitate the examination of more relevant discussions in which the targeted software-related terminology is more likely to be present.

Next, in order to optimize the use of time and resources within the limitations of this project, a series of steps were followed to obtain a smaller sample for the selection of terms. The sampling was carried out in consultation with Professor Pierre-Jerôme Bergeron from the University of Ottawa's Department of Mathematics and Statistics (e-mail conversation, 2 April 2014). A random selection of 10% of the 734 topics with relevant terminology was made in order to achieve the final list of 73 discussion topics that are included in the sample to be manually examined for the discussion of relevant concepts and terms that express those concepts. These

---

[40] For example, questions and discussion relating to the use of disk partitions in Ubuntu appear under several topic titles such as "installation et partition", "réinstallation et partition disque" and "partition disque - assistance". These messages are grouped under the overarching topic title "partition".

[41] A precise word count for the corpus cannot be obtained due to the structural characteristics of the forums. Every posting on the forums includes a large amount of text serving administrative purposes (e.g. author's full name and e-mail address, message subject, links to the previous and the next message posted with the subject titles of these messages, date of posting, etc.). In addition, messages in reply to others include the original message, as in the case of reply e-mails. Moreover, there are many cases where authors copy and paste the original message into their posting, and insert their responses to specific points made in the original message. The vast inconsistency of the amount of administrative and duplicated text as such also prohibits any meaningful estimate or average regarding the number of words in the corpus. Therefore, the number of topic titles and messages posted under titles are considered to be the only meaningful indicators of the size of the corpus.

discussion topics (both topic titles and the postings under the titles) were examined manually for pertinent concepts discussed, with terms used to designate these concepts being recorded. In total, this sample included 180 information-technology related concepts expressed by 252 terms. The time and resources available for this project would not realistically allow for the examination of all 180 concepts in the sample and 252 terms that designate these concepts in the corpus in addition to terms from external terminological resources. In order to limit the workload within the scope and limitations of this project, the four-step selection procedure below was developed and it was followed step by step, handling one overarching concept at a time. This procedure was repeated for a period of two weeks (approximately 80 working hours) allotted for this stage of this project. At the end of the allotted time, a total of 194 terms designating 37 concepts was collected.

1) A random number generator was used to choose a concept from the list of 180 concepts identified within the sample;

2) Terms designating the chosen concept were identified and listed:

   a) Terms designating the chosen concept in the corpus sample were listed;

   b) Terms designating the chosen concept in the following terminology resources were recorded[42]:

      - TERMIUM, the official term bank of the Government of Canada[43],
      - Le Grand dictionnaire terminologique (GDT), the official term bank of the Government of Québec[44], and

[42] In addition to French terms designating the chosen concept, all English terms from the same terminological resources were recorded to account for the possibility of borrowings.
[43] http://www.btb.termiumplus.gc.ca, last accessed 24 June 2016.
[44] http://www.gdt.oqlf.gouv.qc.ca, last accessed 24 June 2016.

- The MS Language Portal, the official terminology database of Microsoft[45].

3) The entire corpus was queried for each term using the monolingual concordancer TextStat[46] and occurrences were counted for the calculation of the relative frequency of each term designating the chosen concept. This included the process of disambiguation, where each occurrence was verified to be an authentic occurrence invoking the relevant overarching concept. The verification of the number of occurrences was carried out twice to reduce the chances of manual error.

4) English terms that were found in the above-mentioned terminology resources without any occurrences in the corpus were excluded from further analysis. In other words, only English terms that are borrowed in French postings in the discussion forums that are part of the corpus are part of the further examination discussed below.[47]

It is worth noting that it is theoretically possible that terms that were not present in the sample or in any of the terminology resources consulted, despite being present in the corpus and referring to the concepts identified initially, were overlooked. However, this is deemed to be within a reasonable margin of error.

3.2.4. Determination of sample size

As this project differs from conventional terminometric research, the sample was created in the course of the two previous stages explained in sub-sections 3.2.2 and 3.2.3. Therefore, details regarding the size of the corpus and the sampling processes that were followed can be found in

---

[45] http://www.microsoft.com/Language, last accessed 24 June 2016.
[46] In light of the languages and the file formats of the documents in the corpus and the functions required for tasks that fall within the scope of this study, this particular tool was deemed sufficient. In addition, the tool is available for download free of charge, and the author has previous experience using the tool. http://neon.niederlandistik.fu-berlin.de/en/textstat, last accessed 24 June 2016.
[47] This led to 60 English terms with no occurrence in the corpus being excluded from further analysis, therefore bringing the total number of terms that were analyzed to 134.

those sub-sections.

3.2.5. Data collection and processing

As a precaution against the disappearance of data from the websites containing the discussion forums in the course of this project, all corpus contents were downloaded and saved in a secure hard drive with the help of the free open-source offline browser, HTTrack[48]. Once the previous stages explained above were complete, the corpus was loaded into the monolingual concordancer TextStat, which partially automates the counting of occurrences of terms and facilitates the disambiguation process.

The corpus was queried for the stem for each term in order to allow the concordancer to retrieve occurrences of plural or inflected forms. The queries were carried out in *case insensitive* mode for the majority of terms to include uppercase and lowercase spellings of the terms being sought. The exception to this concerns queries for acronyms (e.g. *OS* – operating system), where a case insensitive search produces results that contain an extremely high amount of noise (i.e. every occurrence of the character string *OS* is retrieved – e.g. *gros*, *vos*, *infos*, etc.), which would significantly increase the time spent for disambiguation. By searching in *case sensitive* mode, it is possible to retrieve only the characters spelled as they are in the query (whether lowercase or uppercase – uppercase in the case of acronyms), which reduces the amount of noise when it comes to acronyms that are most commonly spelled in uppercase letters. There may be occurrences of acronyms spelled in lowercase letters, especially considering the informal nature of the majority of messages posted on discussion forums. Similarly, there may be other terms appearing in misspelled forms. The queries are formulated using the correctly spelled form(s) of terms as they are recorded in the terminological databases. There may be misspelled occurrences

[48] http://www.httrack.com, last accessed 24 June 2016.

that are excluded; however, this is not considered to be a factor that can affect the results of this study in any significant way.

The immediate context displayed in the list of results is set to 75 characters to the left and to the right of the search string, since this is the approximate amount that fits in a regular wide-screen computer monitor in default screen resolution when the concordancer is used in full-screen mode and allows for a quick examination of the immediate context surrounding the search string. In relatively rare cases when the immediate context is not sufficient for disambiguation, the concordancer allows the user to view a larger portion of the context separately from the list of results, or the entire document from which the context is retrieved. The results can be sorted alphabetically (based on the key word and its variants) or based on the character following or preceding the search string. Alphabetical sorting of concordances is not used in this study. Sorting based on the character preceding or following the search string allows the grouping together of recurring contexts, which facilitates the elimination of redundancy. Since the corpus contains a high number of messages sent in response to others, where the original message and/or the entirety of the conversation might be quoted in each message, the concordance lists contain numerous repetitions. When concordances are sorted based on the character preceding or following the search string, repeated contexts are listed together, and this facilitates identifying the authentic occurrence to be counted only once. This is also facilitated by the fact that quoted portions in reply messages are always marked by characters that can be easily visually identified in the immediate or extended context (e.g. >, <) and excluded from the count. Also excluded are any strings that are part of non-human or non-authentic language such as file paths, URL addresses, programming language, software and brand names and so on, in addition to occurrences that are part of contexts in English (e.g. messages posted in English), since this

study is interested only in term implantation in French. Figure 1 below is a compilation of partial screenshots taken from the occurrences displayed in TextStat, illustrating some of the above-mentioned visual markers circled in red. From top to bottom, the figure shows an authentic occurrence of *curseur* followed by repetitions in reply messages with markers, occurrences of *blogue* as part of a URL, and occurrences of *download* used as a borrowed term, appearing as part of a URL, and in a message composed entirely in English.

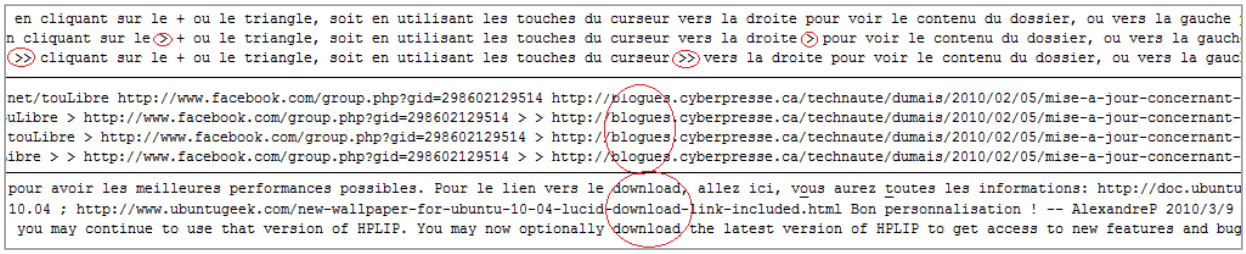

Figure 1 - TextStat screenshot compilation of visual disambiguation markers

The disambiguation stage requires manual effort and is prone to error. Within the scope and limitations of this project, it was not possible to have multiple annotators working on disambiguation, and the time that was available for this task was limited. However, in order to ensure the accuracy of the number of occurrences counted for each term within these limitations, this process was carried out twice without any cases of significant discrepancy between the numbers counted in the two instances. Therefore, it is safe to assume that the margin of manual error in disambiguation is minimal. In addition, the discussion of findings mainly focuses on overall trends rather than specific numbers of occurrences, especially in the case of terms with higher numbers of occurrences, which increases the risk of manual error. Such a minimal margin of error is not likely to seriously affect any conclusion that may be reached.

Based on the number of occurrences of each term after disambiguation, implantation

coefficients were calculated using the formula suggested by Quirion (2003a), seen in Figure 2 below.[49]

$$\frac{\textit{number of occurrences of term T representing the concept N}}{\textit{number of occurrences of terms T + T' representing the concept N}} = \text{implantation coefficient for term T}$$

Figure 2 - Implantation coefficient (Quirion 2003a: 33)

3.2.6. The presentation and discussion of results

The presentation of results in chapter four (Data Analysis) provides an outline of the data obtained in this study through filters with increasing complexity in each section. First, a broad perspective is presented to demonstrate the overall distribution of terms according to term formation patterns and on the basis of average implantation values observed overall and for each pattern. This allows us to establish the basis of the examination that follows, and provides a frame of reference for assessing frequency and implantation within this study. The rest of this chapter moves on to more focused examinations regarding the source of terms, the distribution of terms by term formation patterns, and the distribution of terms by implantation. The application of different filters on the data collected in the study culminate in the juxtaposition of data on the distribution of terms by term formation patterns and the distribution of terms by implantation, with the goal of assessing correlations between the two.

The discussion in chapter five (Discussion of Findings) presents the interpretation of the data

[49]In the case of 3 out of 134 terms, implantation rates were calculated or estimated without counting all occurrences, since the benefits of reducing the time spent on this task and the chances of manual error were deemed superior to the benefit of having an absolute count of all occurrences. If not a single competing term was found in the corpus for a given term, the occurrences of the term were not counted but its implantation coefficient was calculated as 1.00 (this applies to the terms *script* and *clavier*). If an extremely low number of competing terms were found in the corpus for a given term and the occurrences of this term were visibly far more numerous, its implantation coefficient was estimated as 0.99 (this applies to the term *bouton*).

in light of the research question, hypothesis and objectives of this study, and on the basis of the examination presented in chapter four. Within the previously acknowledged limitations of the study, correlations between term implantation and term formation patterns constitute the focus of discussion. Observations that apply to specific term formation patterns, and emergent findings that were not part of the main objectives of the study are presented with indications of their implications for future research.

4. DATA ANALYSIS

This chapter presents the quantitative data gathered from the corpus. Since term formation patterns and term implantation are the two essential indicators identified for the purposes of this study, data analysis is focused on these indicators with alternating priorities assigned to each of the factors that are part of the study in the sections below.

Although implantation coefficients allow us to assess the implantation of each term designating a concept within the boundaries of that individual concept, it is difficult to evaluate them on their own. Quirion's (2003b: 29, 30) overview of terminometric research literature demonstrates the variety of strategies suggested for determining the success of implantation, ranging from setting somewhat arbitrary thresholds of relative frequency to looking for usage in the majority of instances. This leads the author to conclude that, ultimately, implantation is relative. Therefore, the approach introduced in this study relies on establishing an internal point of reference in order to refrain from setting arbitrary definitions for assessing levels of implantation. Instead, this approach allows for an assessment relative to the context of this study (i.e. within the same corpus, within a particular group of users, within a defined period of time, etc.). This is done by calculating the average IC value of the 134 terms analyzed. Within this study, this average can serve as an objective point of reference representing the distribution of terms by implantation, providing a useful, novel and consistent basis for the discussion of levels of implantation. The average IC value for the 134 terms analyzed in this study is 0.28. This means that the average term is used to designate the relevant overarching concept in 28% of instances where the concept is evoked. This value will hereafter be referred to as the *reference average IC.*

In order to facilitate the observation of potential trends or tendencies correlated with term

implantation, four ranges of implantation are used to group terms based on their IC levels. Hereafter, *terms with no implantation* refers to terms with an IC of 0; *terms with lower implantation* refers to terms with an IC higher than 0 and lower than or equal to the reference average IC of 0.28; *terms with higher implantation* refers to terms with an IC higher than 0.28 and lower than 1; and *fully implanted terms* refers to terms with an IC of 1 (see Table 2 below).

IC Range	Range Title
IC = 0	Terms with no implantation
0 < IC ≤ 0.28	Terms with lower implantation
0.28 < IC < 1	Terms with higher implantation
IC = 1	Fully implanted terms

Table 2 - Ranges of implantation

The first section (4.1) outlines general observations that apply to all of the collected data from a broad perspective regarding the distribution of terms exhibiting each term formation pattern and the average IC values observed for each pattern. Sections 4.2 and 4.3 examine the distribution of terms based on the sources in which they were observed, and based on the term formation patterns they exhibit, respectively. Section 4.4 outlines the distribution of terms in the four above-mentioned ranges of implantation. Finally, section 4.5 presents a synthesis of term distribution on the basis of term formation patterns and implantation with the goal of assessing potential correlations between these two elements.

4.1. Overall term distribution

This sub-section outlines the overall distribution of the 134 terms analyzed in this study based on the percentage of terms exhibiting each formation pattern and the average IC value calculated for

the group of terms exhibiting each pattern, as illustrated in Table 3 below. It is worth remembering here that term formation patterns are not mutually exclusive. In other words, there are many terms that exhibit multiple formation patterns. A complete list of terms, with the term formation patterns they exhibit, their sources and their IC values can be found in Appendix 3.

Term formation pattern	**Number of terms**	**% of terms that exhibit pattern**	**Average IC**
Adapted borrowing	6	4%	0.34
Loan translation	3	2%	0.27
Direct borrowing	24	18%	0.24
Derivation/Affixation	61	46%	0.21
Compounding	71	53%	0.20
Use of existing resources	56	42%	0.18
New creation	17	13%	0.16
Compression	10	7%	0.16
ALL	***134***	***N/A***	***0.28***

Table 3 - Overall distribution of terms and IC levels

First of all, it should be noted that the total frequency of term formation patterns does not add up to 100% and this is due to the fact that many terms exhibit multiple formation patterns. The table is sorted by the average IC value from the highest to the lowest. The only pattern that has an above-average IC level is adapted borrowing, and borrowings have the three highest average IC levels. This is in part due to the fact that borrowings include only English terms found in the corpus. In other words, there are no borrowed terms that have an IC level of zero, and this results in higher averages for these patterns.

The fact that all but one of the patterns have below-average IC values reflects the high level of competition among terms designating the same concept. In other words, the low IC

averages reflect the lack of standardization that is frequently identified in the literature as a phenomenon commonly observed in rapidly developing fields related to technology (cf. Sager 1990; Corbolante and Irmler 1997; Lepouras and Weir 1999).

Aside from providing a general idea, overall distribution and average values provide limited information. The following sections examine the data more closely – namely based on sources, patterns, IC levels, and a juxtaposition of patterns and IC levels – for the purpose of reaching a better understanding of potential correlations between term formation patterns and term implantation.

4.2. Distribution by source

As explained in the previous chapter (sub-section 3.2.3), 194 terms from the corpus and from external terminological resources (TERMIUM, the GDT and the MS Language Portal) were initially identified for examination, prior to the exclusion of 60 English terms without occurrences in the corpus. This section presents a closer look at term implantation levels in relation to where the terms are found. Implantation levels of terms found in the corpus sample, TERMIUM, the GDT and the MS Language Portal are examined for potential correlations between sources and implantation levels. This examination can also be useful in understanding the coverage of terms by each terminological resource in this particular context.

Table 4 below shows the distribution of terms according to the source(s) in which they were observed. It is worth noting that most terms are observed in more than one source. The sources are presented in order of the number of concepts covered, thus starting with the sample as the origin of all the concepts analyzed in this study (as explained in the previous chapter, in sub-section 3.2.3).

Source	Total concepts / terms
Sample	37 / 55
MS Language Portal	34 / 47
TERMIUM	33 / 141
GDT	32 / 113
Total # of terms	*37 / 194*

Table 4 - Distribution of concepts and terms by source

The MS Language Portal follows the sample with the second highest coverage of concepts, followed by TERMIUM and the GDT respectively. However, the highest number of terms are found in TERMIUM, followed by the GDT, the corpus sample and finally the MS Language Portal. The differences in the coverage of terms and concepts in each source, and the implantation of terms from each source are further discussed in the sub-sections below.

4.2.1. Sample

Terms observed in the sample are those that were identified in the initial stages of the study from within the sample taken from the corpus (as explained previously in sub-section 3.2.3). They constitute 28% of all terms sought in the corpus (55 out of 194), covering all 37 concepts with an average of 1.5 terms per concept. The low number of terms per concept indicates a high rate of implantation. This is reflected in the average IC of the terms in the sample, which is notably higher than the reference average IC of 0.28, at 0.58. This is due to the fact that this group of terms is included in the study on the basis of their presence in the corpus, which evidently leads to higher average IC levels as there are no terms in this group with zero implantation. Out of the 55 terms observed in the sample, 16 terms (29%) exhibit IC levels below average. The majority of these terms (39 terms, approximately 71%) have above-average IC levels, including nine

terms that are fully implanted (see Figure 3 below[50]).

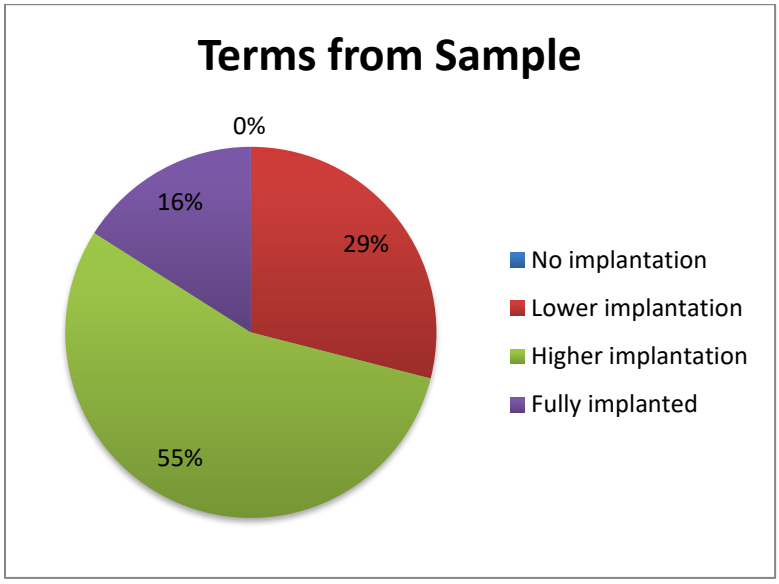

Figure 3 - Distribution of terms from the corpus sample by implantation

4.2.2. The MS Language Portal

Although the coverage of the number of terms (47 out of 194, as seen in Table 4 above) may appear low, the MS Language Portal compares favourably to the other terminological databases on the basis of the coverage of concepts and the implantation levels of the terms covered. This terminological resource has the highest coverage of concepts, with 34 out of the 37 concepts examined. These 34 concepts are designated by 47 terms, with an average of 1.4 terms per concept. Terms found in this resource have an average IC level of 0.56, which is the highest average IC observed among terms recorded in the three terminological databases consulted.

Of the 47 terms found in the MS Language Portal, eight terms (17%) are fully implanted.

[50] The percentages in the pie charts presented in this document are rounded up, so totals may exceed 100%.

This is the highest rate of fully implanted terms among all sources. The number of terms that have above-average IC levels (excluding fully implanted terms) is 21 (44%), and 11 terms (23%) are found in the corpus with below-average IC levels. The number of terms found in this resource that do not have any representation in the corpus is seven. This is the lowest rate of terms with no implantation among all three terminology databases. Figure 4 illustrates the distribution of terms found in this resource by implantation.

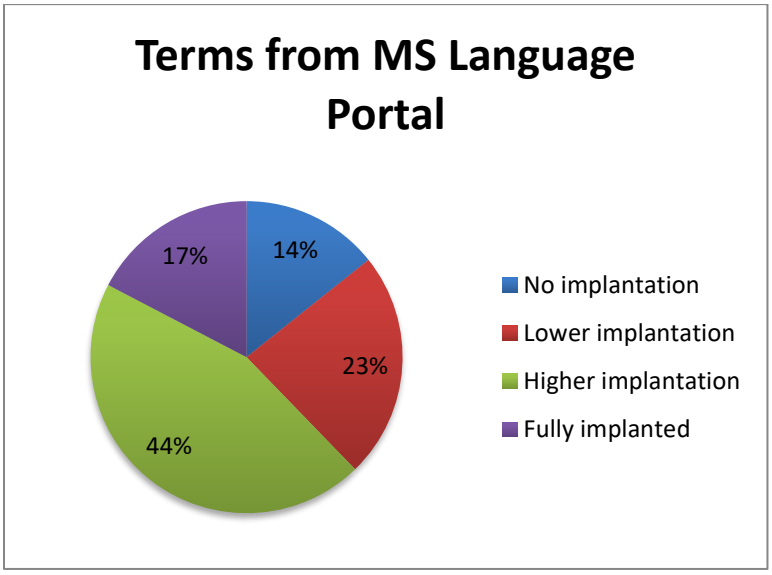

Figure 4 - Distribution of terms from the MS Language Portal by implantation

The presence of the above-mentioned positive indicators that lead to a favourable comparison regarding term implantation for the MS Language Portal over the other terminological databases may be due to several factors. One of these factors may have to do with the fact that this resource is specifically prepared for the information technology domain by Microsoft, which is one of the most prominent producers of content within this domain as a major producer of software and operating systems. It is reasonable to assume that this prominence would result in a high level of

influence over the terminology used in this domain. In addition, the main purpose of the MS Language Portal is to provide a concise and reliable source of standardized terminology for Microsoft products, minimizing the number of terms that express a specific concept for the sake of clarity and consistency within and among software products. Meanwhile, terminological resources such as TERMIUM and the GDT aim to provide more extensive coverage of terminology in every subject field, covering as many competing terms as possible so as to provide users with terminological information that is as exhaustive as possible. Finally, it is worth mentioning that Microsoft has launched and maintained a community forum for terminology, the Microsoft Terminology Community Forum, to "enable community members to develop, discuss, and approve new terminology for Microsoft products" (DePalma and Kelly 2011: 289), which has since been integrated with the MS Language Portal. The terminology database accessible through the Portal includes reviewed and finalized input from the forum. This process facilitates the promotion of terms "that reflect desires of future product users" (Karsch 2014: 300). The collaborative aspect of MS Language Portal in comparison with terminology databases maintained by governmental institutions may, therefore, be a factor in the high amount of correspondence between the terminology used in the corpus and the terminology stored in this resource.

Although this project does not involve a comparison of terminology use in Microsoft products with that in open-source software, or of terminology use by Microsoft users with that of open-source software users, the information presented above suggests that terminology use in the domain of information technology is most likely influenced to some extent by the terminology disseminated by the major producers of content within this domain.

4.2.3. TERMIUM

Of all the terms initially identified for examination, 73% (141 out of 194) were found in TERMIUM, which means that this terminological resource covers the highest number of terms among all sources. However, the high number of terms stored for each concept (141 terms designating 33 concepts – over four terms per concept on average) and the fact that the majority of these terms are not found in the corpus manifest themselves in a low average IC of 0.19. Out of the 141 terms found in TERMIUM, 78 terms (55%) are not found in the corpus, 31 terms (22%) are found in the corpus with below-average IC levels, and 32 terms (22%) have above-average levels, including six terms that are fully implanted (see Figure 5 below).

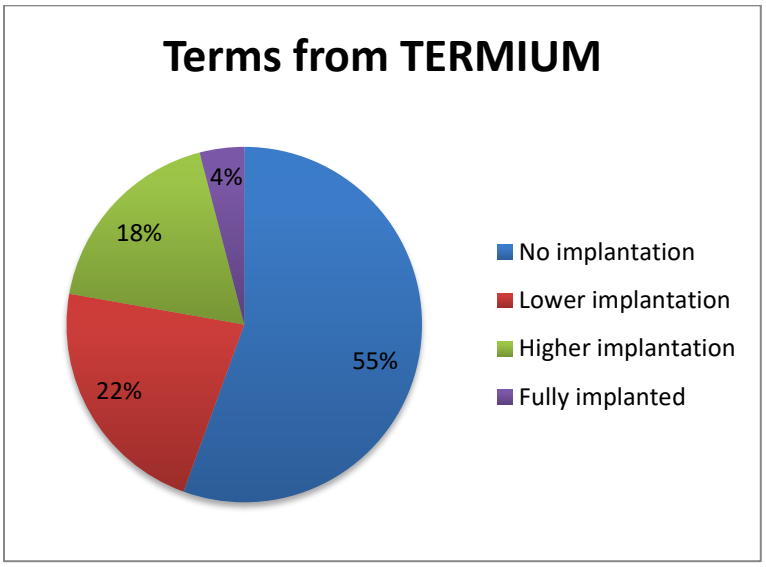

Figure 5 - Distribution of terms from TERMIUM by implantation

The findings demonstrated above indicate that TERMIUM has a wide coverage of terminology, although a significant portion of the terminology covered in this particular context was not found in the corpus. It is worth mentioning here that terms stored in TERMIUM were included in this study regardless of any usage labels or recommendations that may be stored in term records to

indicate "currency of use, locality, social context, etc."[51]. The wide coverage of terminology in large institutional terminology databases such as TERMIUM and the GDT (examined below) serves an important purpose within this study regardless of implantation, as it reduces the chances of overlooking terms designating one of the selected concepts for terminometric analysis.

4.2.4. The GDT

The number of terms observed in the GDT is 113, which constitutes 58% of the 194 terms initially identified for analysis. This is the second highest number of terms covered in the terminological databases consulted in this study. These terms designate 32 concepts, which means three to four terms are stored per concept, on average. The average IC level of terms found in the GDT is 0.23. Of the 113 terms found in this terminological database, 60 (53%) are not found in the corpus, 23 (20%) are found in the corpus with below-average IC levels, 30 (26%) have above-average IC levels, including six terms (5%) that are fully implanted (see Figure 6 below).

[51] http://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=&index=alt&srchtxt=USAGE%20LABEL, last accessed 24 June 2016.

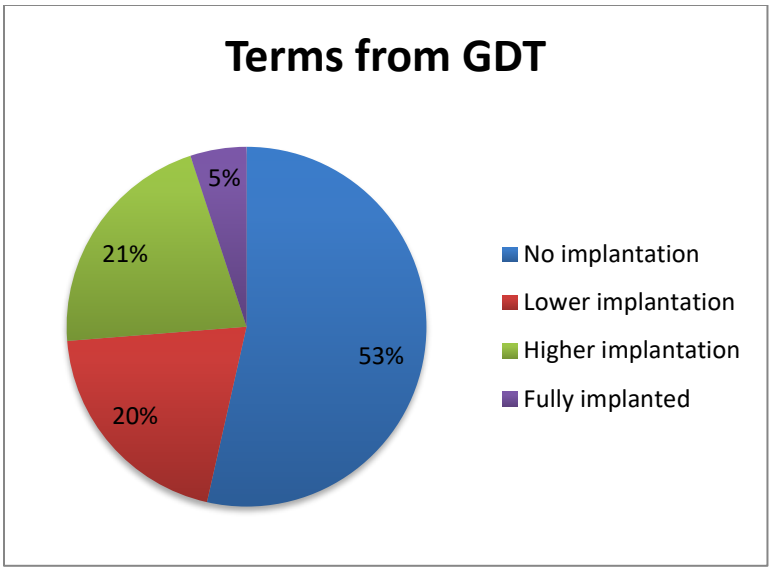

Figure 6 - Distribution of terms from the GDT by implantation

4.3. Distribution by pattern

This section examines the distribution of terms by term formation pattern. Before we go into a detailed examination of each individual pattern on the basis of implantation levels of terms that exhibit the pattern, Figure 7 below shows the distribution of the 134 terms that underwent terminometric analysis in this study by the term formation patterns they exhibit.

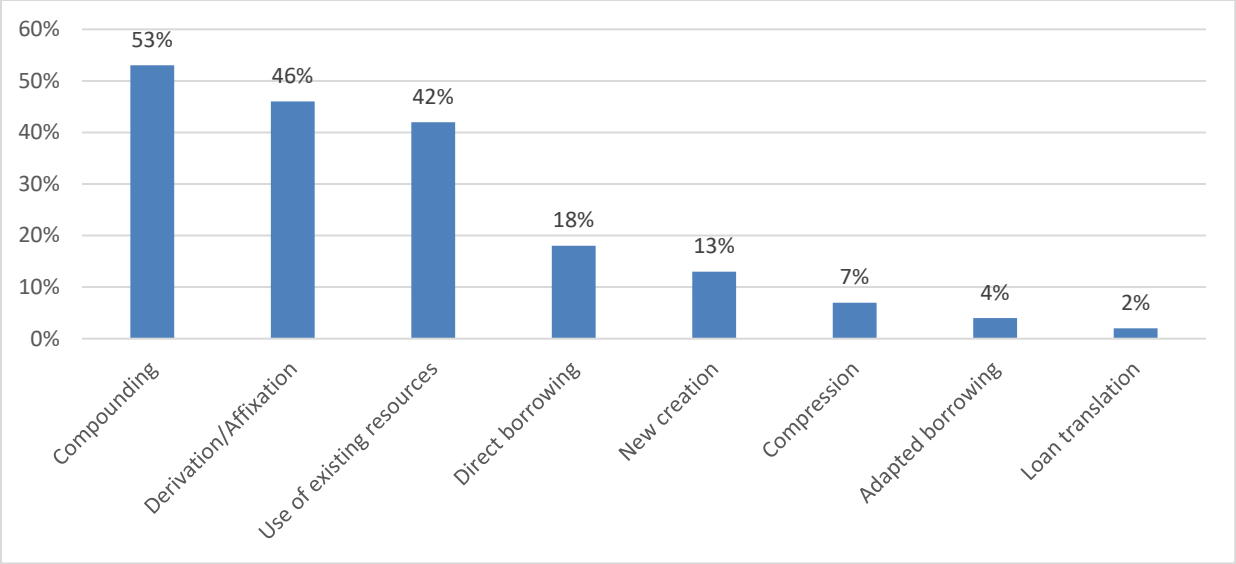

Figure 7 – Overall frequency of term formation patterns

This overall distribution serves as a basis for further examination, which follows below. The three most frequently observed patterns that stand out from the others by a margin of over 20% are, respectively, 1) compounding, 2) derivation/affixation and 3) the use of existing resources. The sub-sections below examine each pattern and the implantation rates of terms exhibiting each pattern in more detail, from the most frequently observed pattern to the least.

4.3.1. Compounding

Of the 134 terms analyzed, 71 are formed through compounding. Thus, compounding is the most commonly observed term formation pattern within this study. Some examples for terms that exhibit this pattern include the following: *fournisseur internet, carte-mère, bureau à distance, gestionnaire de mise à jour.* Designating the concepts *system monitor* and *text processing*, respectively, the terms *système d'auto-contrôle* and *traitement des mots* are compound terms with no implantation, while *moniteur de système* and *traitement de texte* are among fully implanted compound terms.

The average IC observed within this group is 0.20. Figure 8 demonstrates the distribution of terms exhibiting this pattern on the basis of their IC level in four ranges of implantation: 1) terms with no implantation, 2) terms with lower implantation, 3) terms with higher implantation, and 4) fully implanted terms.

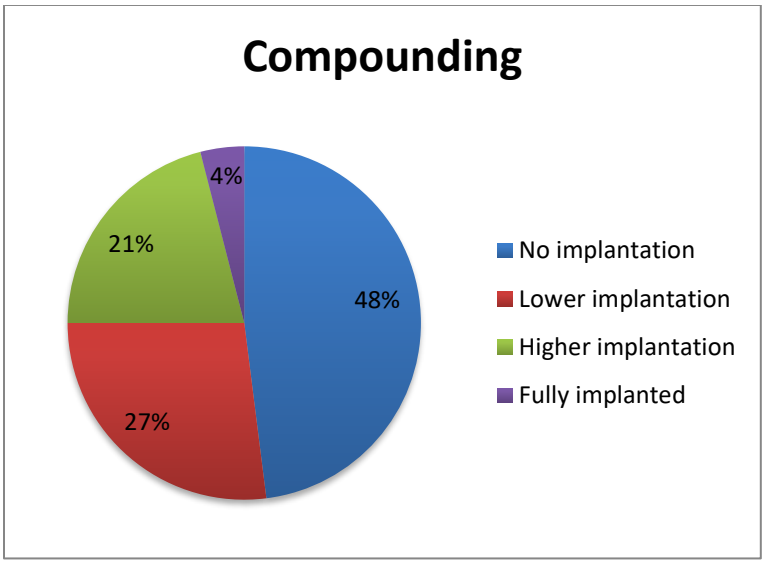

Figure 8 - Distribution of compound terms by implantation

Close to half (48%) of compound terms are not found in the corpus. More than a quarter (27%) of these terms are observed with lower implantation rates, and 25% are observed with above-average implantation, including 4% that consists of fully implanted terms.

4.3.2. Derivation/affixation

As the second most frequently observed term formation pattern, derivation/affixation covers 61 terms out of 134. Some examples for derived/affixed terms are: *affichage, processeur, programmation, téléchargement, utilisateur/-trice.* The terms *interfonctionnement* and *démontable* are examples of derived/affixed terms with no implantation, while *interopérabilité*

and *amovible* are fully implanted derived/affixed terms designating the concepts *interoperability* and *removable,* respectively.

The average IC observed within this group is 0.21. Figure 9 below demonstrates the distribution of terms exhibiting this pattern on the basis of four ranges of implantation.

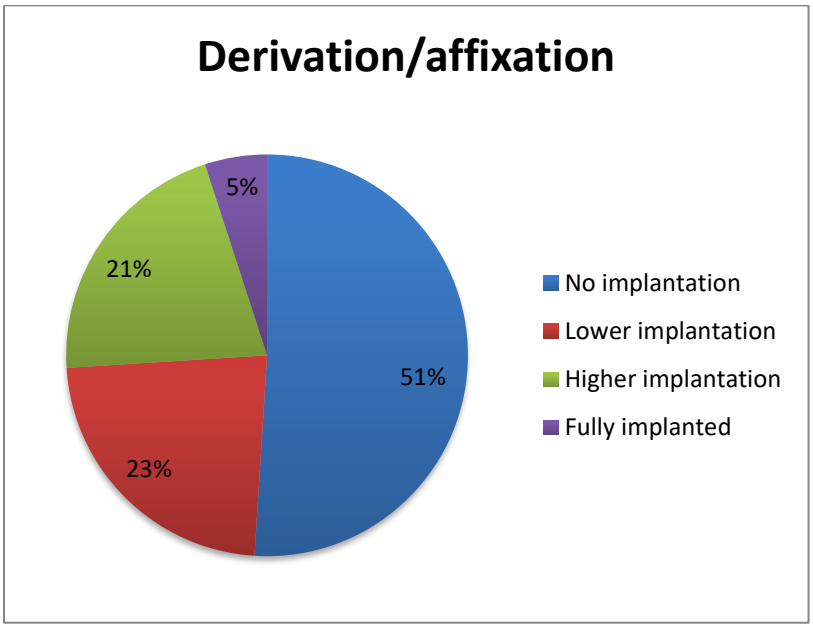

Figure 9 - Distribution of derived/affixed terms by implantation

A vast majority (74%) of terms formed through derivation/affixation have below-average implantation, with over half (51%) not found in the corpus, and 23% figuring among terms with lower implantation. The rate of terms exhibiting this pattern with above-average IC levels is 26%, which includes 5% that consist of fully implanted terms. Therefore, this pattern follows the use of existing resources with the second highest rate of fully implanted terms.

4.3.3. Use of existing resources

The use of existing resources is the third most frequently observed pattern among the 134 terms analyzed, with 56 terms exhibiting this pattern. Examples of terms that exhibit this pattern

include *bouton, léger, mémoire,* and *script.* The terms *en direct* and *scénario* figure among terms exhibiting this pattern with no implantation, and the terms *curseur* and *clavier* are examples of fully implanted terms formed through the use of existing resources.

The average IC observed in this group is 0.18. Figure 10 demonstrates the distribution of terms exhibiting this pattern on the basis of their IC level in four ranges of implantation.

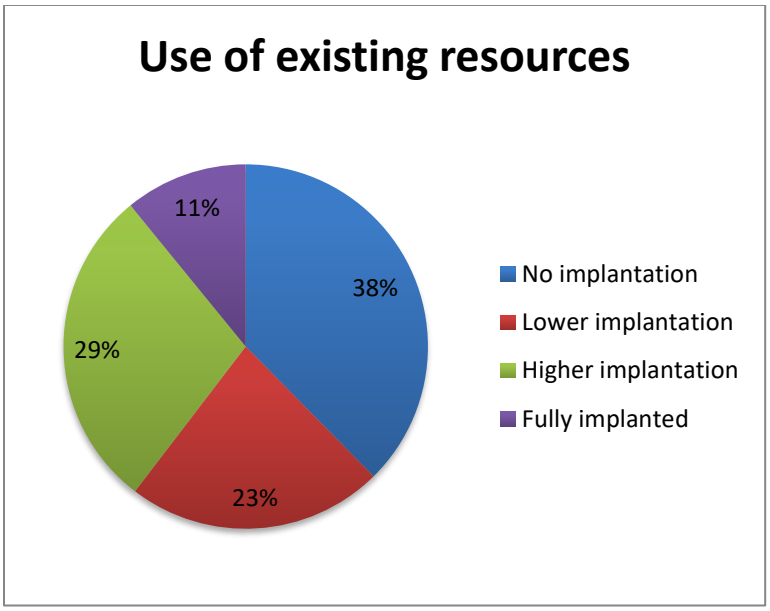

Figure 10 - Distribution of terms formed through the use of existing resources by implantation

A little over one third of the terms formed through the use of existing resources (38%) are not observed in the corpus, and 23% of these terms are observed within the lower implantation range. The rate of terms exhibiting this pattern with above-average implantation is 40%, including 11% that are fully implanted. Accordingly, this pattern includes the lowest rate of terms with no implantation as well as the highest rates of fully implanted terms and of terms with above-average implantation among all term formation patterns.

4.3.4. Direct borrowing

Of the 84 English terms that designate the 37 concepts identified for examination in the initial stages of this study – as found in the corpus and in the terminology resources consulted (see sub-section 3.2.3) – 60 terms have no occurrence in the corpus. Hence, the remaining 24 terms are examined as direct borrowings. These terms include *blog, CPU, default, light, login, motherboard, OS, touchpad, trackpad, spam,* and *update manager.* The terms *online* and *user* are among direct borrowings with the lowest implantation rates (IC=0.01), and *multiboot* is the only fully implanted direct borrowing.

The average IC observed within this group is 0.24. Figure 11 demonstrates the distribution of terms exhibiting this pattern on the basis of their IC level in four ranges of implantation.

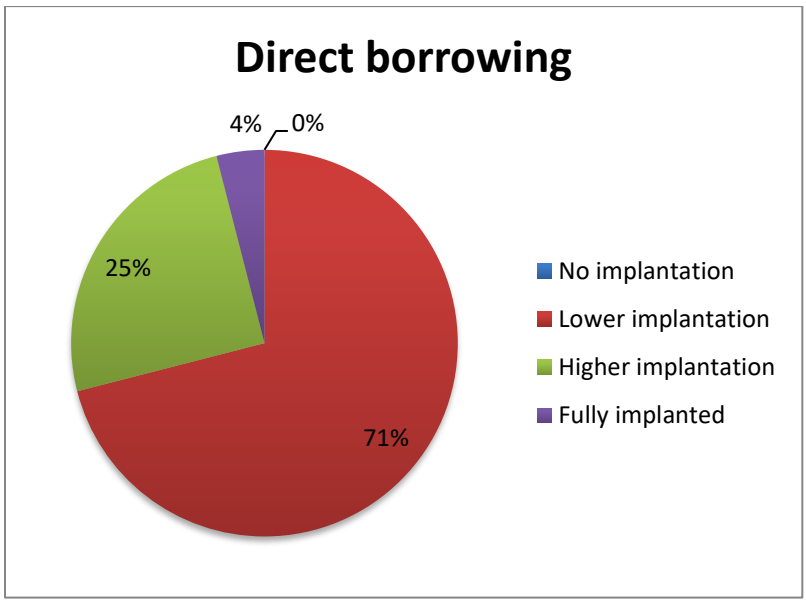

Figure 11 - Distribution of direct borrowings by implantation

As seen in the figure, 71% of direct borrowings fall within the lower implantation range. It is

worth reminding here that there are no direct borrowings with zero implantation in this study, since only English terms observed in the corpus are considered within this pattern. The remaining 29% is observed with above-average implantation.

4.3.5. New creation

Of the 134 terms examined, 17 are identified as new creations. Examples of new creations include *courriel, logithèque, mél, pourriel.* While there are no fully implanted terms exhibiting this pattern, the terms *télécharger. téléchargement* (IC=0.87) and *courriel* (IC=0.80) have the highest levels of implantation among new creations, and *mél* and *méga-ordinateur* figure among terms exhibiting this pattern with no implantation.

The average IC observed within this group is 0.16. New creations (and compressed terms, sharing the same average IC) thus have the lowest average IC among all patterns. Figure 12 demonstrates the distribution of new creations on the basis of their IC level in four ranges of implantation.

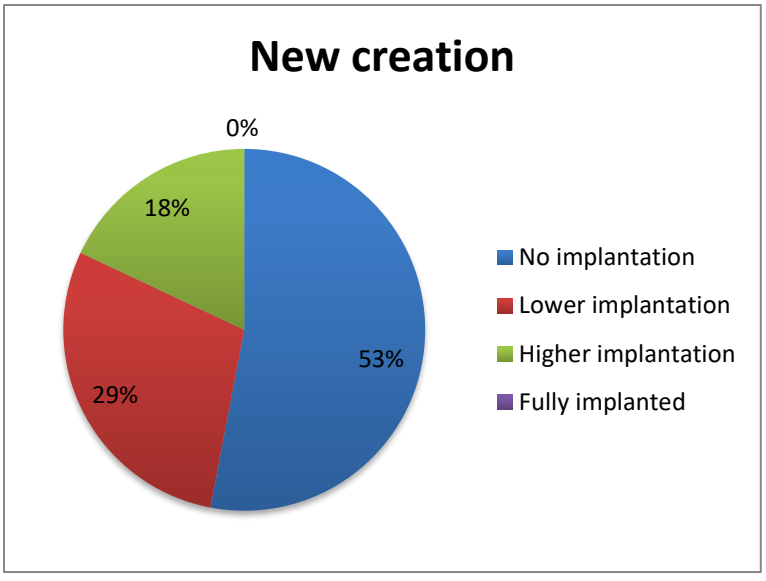

Figure 12 - Distribution of new creations by implantation

A vast majority (82%) of new creations have below-average implantation, with over half (53%) not found in the corpus and 29% found within the lower implantation range. New creations with higher implantation constitute 18% of terms exhibiting this pattern.

4.3.6. Compression

This pattern is observed in 10 out of the 134 terms. Terms that exhibit this pattern consist of abbreviations and acronyms, and of compressed compound terms where one or more of the components are clipped. For instance, the acronym *CÉ* (for *courrier électronique*) is an example of a compressed term with no implantation, while the clipped form *courriel* is the most implanted compressed term (IC=0.80).

Terms formed through compression share the lowest average IC among all patterns with new creations at 0.16. Among the 10 compressed terms, five are not found in the corpus, three are found in the lower implantation range, and two in the higher implantation range. None of the compressed terms are fully implanted.[52]

4.3.7. Adapted borrowing

All six adapted borrowings observed in this study are terms that are variously adapted from English into French. Three of these terms are observed in the lower implantation range (*carnet web,* IC=0.01; *suite office,* IC=0.06; and *suite Office,* IC=0.10[53]), and the other three are

[52] No graphs are presented for the patterns of compression, adapted borrowing and loan translation, examined in sub-sections 4.3.6, 4.3.7 and 4.3.8 respectively, due to the low number of terms exhibiting these patterns.

[53] The terms *suite office* and *suite Office* are examined separately due to the difference in the use of capitalization. The term *suite Office* is most frequently used to refer specifically to Microsoft Office – hence the capitalization, marking the use of *Office* as part of the brand name. However, there were also several occurrences of *suite Office* in reference to office suites in general, indicating a misuse of capitalization. Similarly, there were occurrences of *suite office* in reference to Microsoft Office. Thus, the only occurrences that were taken into account for the calculation of IC values of terms designating the concept *office suite* were those designating the generic meaning, as identified

observed in the higher implantation range (*blogue,* IC=0.45; *librairie,* IC=0.45; and *page web,* IC=0.95).

The average IC observed within this group is 0.34, which is the highest average IC among all patterns. However, this is not necessarily a positive indicator for this pattern, as the individual IC rates for each term exhibit a very scattered distribution across the implantation scale. The high average IC can also be partly explained by the low number of terms exhibiting this pattern.

4.3.8. Loan translation

Loan translation is the least frequently observed term formation pattern, with only three terms among 134 exhibiting this pattern. Two of these terms have lower implantation levels (*bibliothèque de logiciels,* IC=0.01; and *bureau distant,* IC=0.10), and one is among terms with higher implantation (*carte-mère,* IC=0.69).

The average IC observed within this group is 0.27, the second highest average IC among all patterns. However, similar to the other types of borrowing, this high average IC is not a strong indicator, mainly due to the low number of terms exhibiting this pattern and to the scattered distribution of individual IC values.

4.4. Distribution by implantation

This section presents the distribution of terms primarily on the basis of four ranges of implantation representing terms with no implantation, lower implantation, higher implantation

---

through an examination of the context. For example, the following occurrence was excluded as it refers specifically to Microsoft Office and is not a competing term for the concept *office suite*:
"Il n'est même pas question de Linux ici et de changement de système d'exploitation, mais seulement de changement sur un poste de travail qui peut rester en XP (ou vista s'ils sont très masochistes) et de remplacer la *suite office* par la suite bureautique libre OpenOffice.org par exemple..."
The following occurrence, on the other hand, was included as it refers to the concept *office suite*:
"De plus, il y a google aussi qui mange du terrain sur Microsoft en présentant une *suite office* sur Internet."

and full implantation, and how frequently each term formation pattern is observed within each range. Figure 13 below demonstrates the distribution of 134 terms in the above-mentioned ranges of implantation.

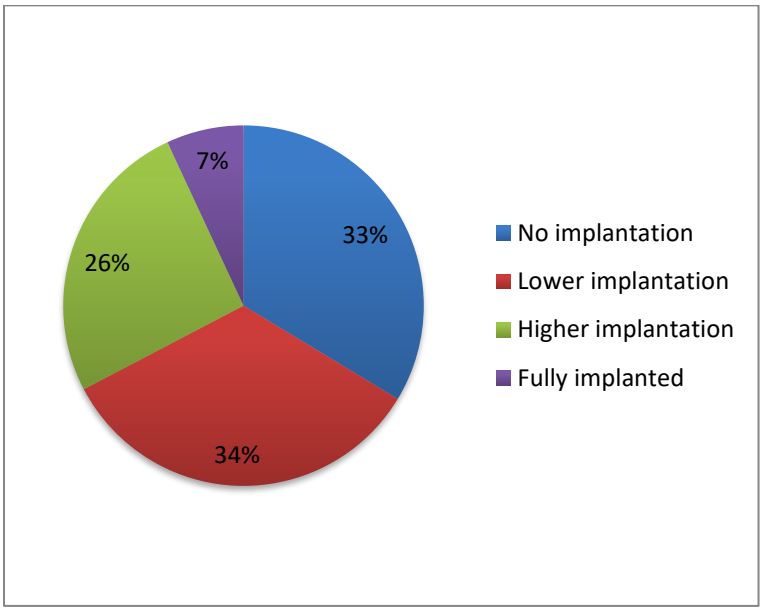

Figure 13 - Distribution of terms by implantation

The majority of the 134 terms (approximately 67%) have below-average IC levels, including approximately 33% with no implantation. This is mainly due to the fact that a significant number of the terms studied come from the terminological databases consulted as part of this study, and do not have occurrences in the corpus. Of the 45 terms with no implantation, 22 were found only in TERMIUM, 11 were found only in the GDT, 2 were found only in the MS Language Portal, 9 were found in TERMIUM and the GDT, and 1 was found in TERMIUM, the GDT and the MS Language Portal (see Figure 14 below).

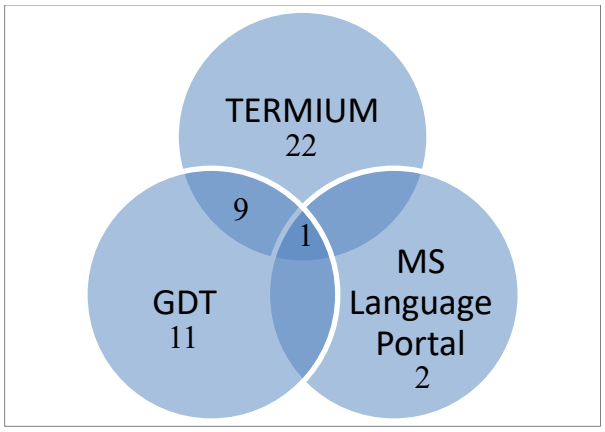

Figure 14 - Number of terms with no implantation by source

The terms with above-average implantation cover approximately 33% of the total and include 7% consisting of fully implanted terms. Thus, terms with no implantation, those with below-average implantation and those with above-average implantation each constitute about a third of all terms.

Before a more detailed examination of corpus data in relation to specific term formation patterns, let us examine the data according to the following four groups: 1) terms with no implantation, 2) terms with lower implantation, 3) terms with higher implantation, and 4) fully implanted terms. For reference, term formation pattern frequencies within each range will be compared with their respective frequency observed overall (as illustrated previously in Figure 7, p.81).

4.4.1. Terms with no implantation

This sub-section examines terms with no implantation on the basis of the frequency of each term formation pattern observed within this group in comparison to the frequency overall, as illustrated in Figure 15 below.

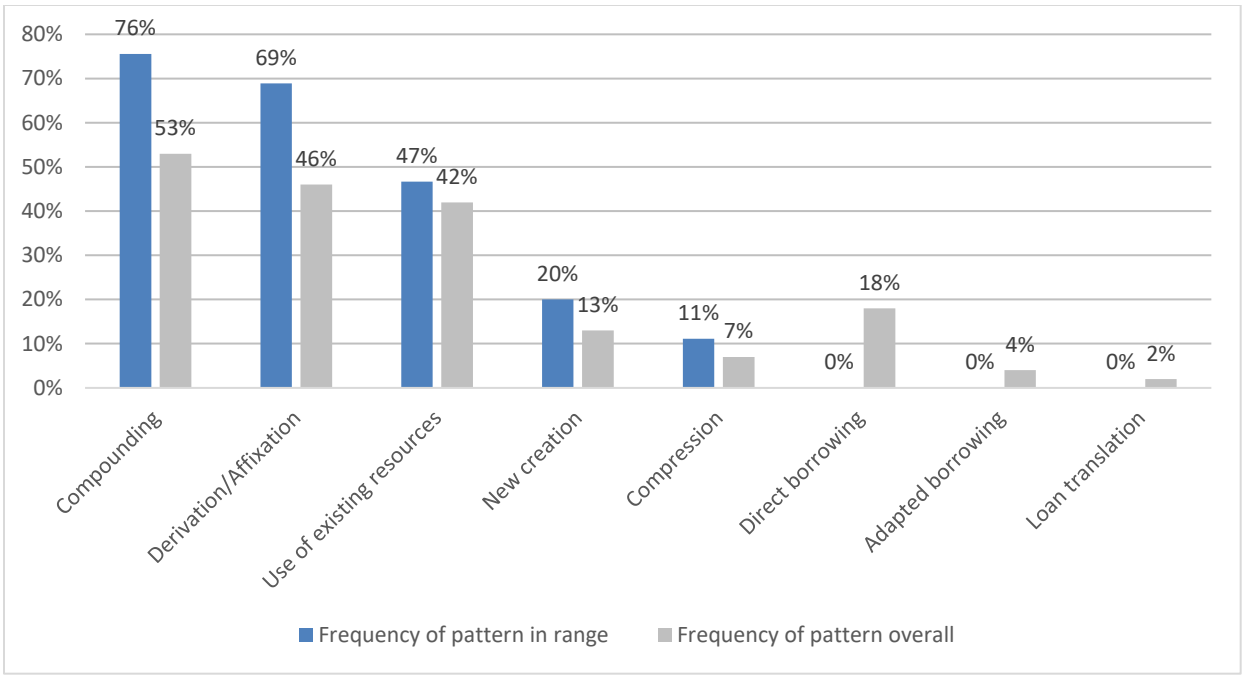

Figure 15 - Frequency of term formation patterns among terms with no implantation

We can see in this figure, for example, that the compounding pattern is observed in 76% of terms with no implantation, whereas it is observed in 53% of the 134 terms. We can observe that, at the lowest end of the horizontal axis, the three types of borrowings are not present among terms with no implantation. The majority of patterns exhibit higher frequencies among terms with no implantation compared to their overall frequencies. This coincides with the low average rates of implantation and high numbers of competing terms observed in general.

Among patterns observed in terms with no implantation, compression and new creation appear with low frequencies that are close to their frequencies in the overall distribution. Next, use of existing resources appears as a pattern that is observed in close to half (47%) of the terms with no implantation, marking a notable increase of close to 30% compared to the patterns with lower frequencies in this range of implantation. This is followed by another sharp increase of more than 20%, with the derivation/affixation and compounding patterns observed in 69% and

76%, respectively, of terms with no implantation.

4.4.2. Terms with lower implantation

This sub-section examines terms with lower implantation on the basis of the frequency of each term formation pattern observed within this group, as illustrated in Figure 16 below.

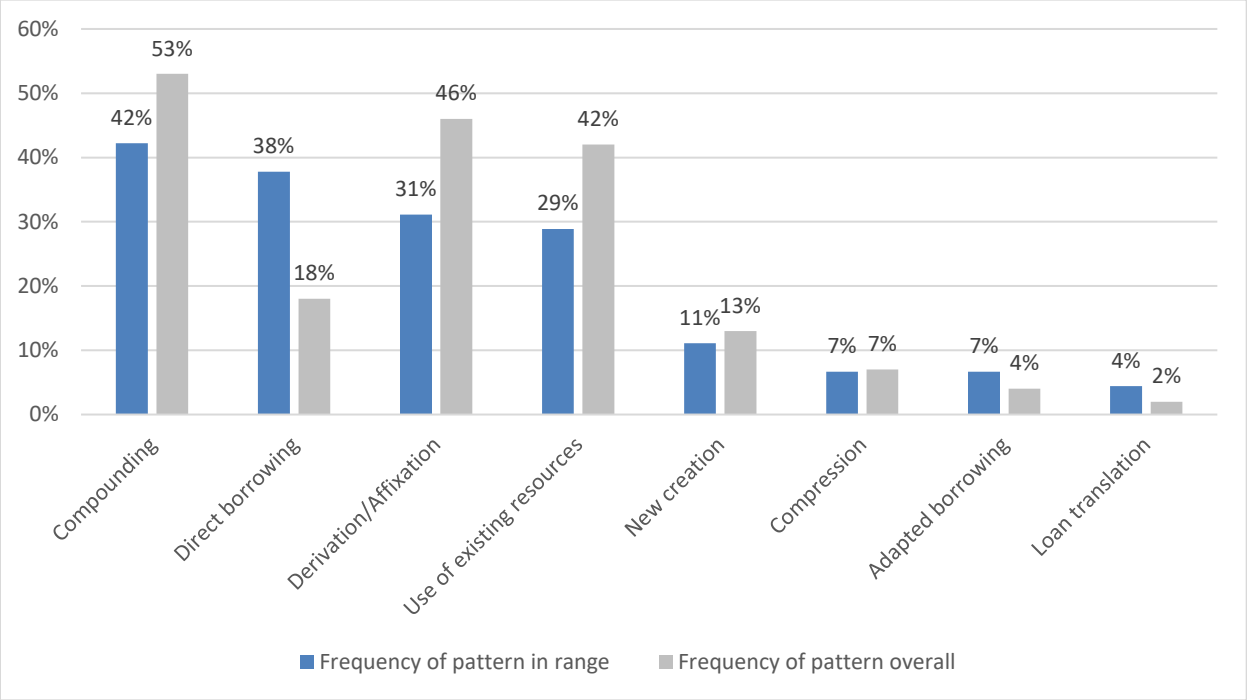

Figure 16 - Frequency of term formation patterns among terms with lower implantation

Although the frequency of most patterns observed in terms with lower implantation is quite similar to terms with no implantation, the frequency of direct borrowings within this range is in contrast with both the overall frequency of the pattern, and its frequency within the other ranges of implantation. In the overall distribution of terms by term formation pattern, and within every other range of implantation, the three most frequently observed patterns are compounding, derivation/affixation and the use of existing resources. In the lower implantation range, however, direct borrowings are the second most frequently observed pattern – 38% of terms within this

range are direct borrowings. Consequently, the lower implantation range is where direct borrowings are most frequently observed.

Loan translations and adapted borrowings both have low frequencies overall, which explains the lower frequencies they demonstrate in every range of implantation, similar to compressed terms and new creations. From the least to the most frequently observed patterns within this range, the marked increase in frequency following 11% of terms exhibiting new creations to 29% exhibiting the use of existing resources is also similar to how term formation patterns compare in frequency overall.

4.4.3. Terms with higher implantation

This sub-section examines terms with higher implantation. There are 35 terms within this range, out of 134 terms (approximately 26%) examined in this study. Figure 17 below shows the distribution of terms with higher implantation based on the frequency of term formation patterns observed within this range.

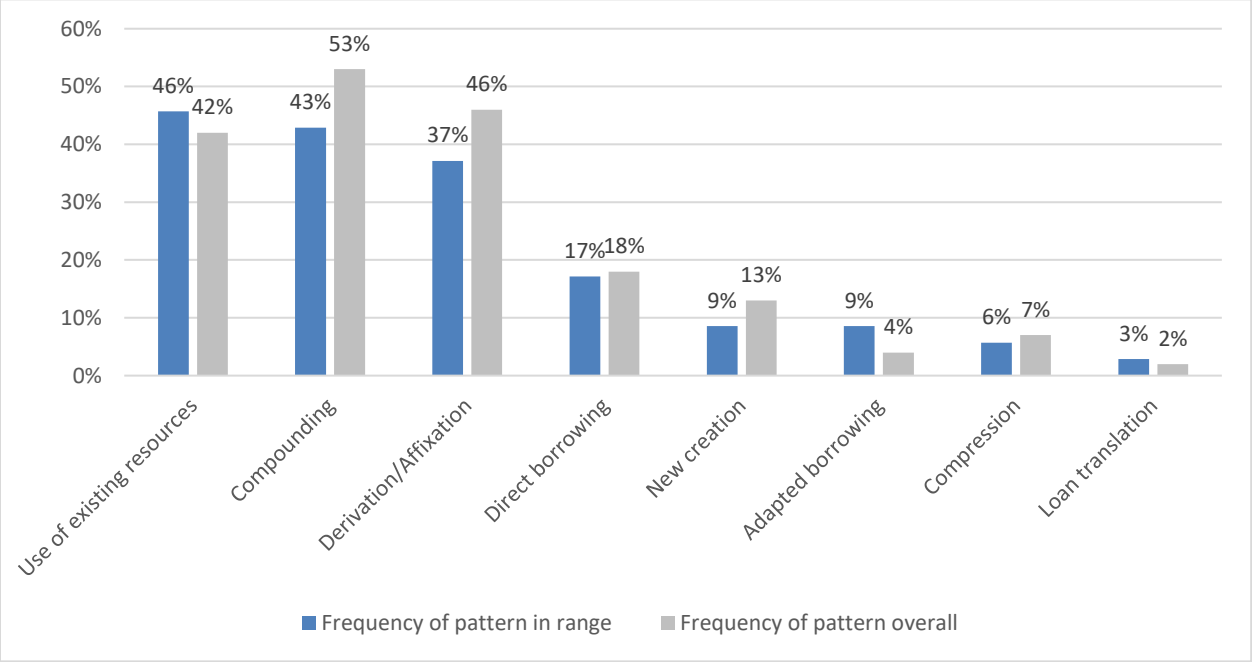

Figure 17 - Frequency of term formation patterns among terms with higher implantation

This distribution shows several similarities with the overall distribution of terms by term formation pattern. The three most common patterns within this group are the same as those seen in the overall distribution: 1) use of existing resources, 2) compounding and 3) derivation/affixation. It is worth noting, however, that the use of existing resources is the most frequently observed pattern among terms with higher implantation and fully implanted terms (explained in the next sub-section). The lower frequencies observed in most of the remaining patterns within this range are relatively consistent with their overall frequency as well. The only other difference in the order of frequency is the reversal of adapted borrowings and compressed terms as the second and third least frequently observed patterns.

4.4.4. Fully implanted terms

Of the 134 terms that are examined, nine are fully implanted. In other words, these terms have an IC of 1.00, which means none of the competing terms designating the same concept as these terms were found in the corpus.

Six fully implanted terms are formed through the use of existing resources, three are formed through derivation/affixation, three of them are compound terms, and one is a direct borrowing. Accordingly, the use of existing resources is the most frequently observed pattern among fully implanted terms. In addition, representing six out of nine terms (or approximately 67%), the use of existing resources has the highest rate of occurrence among fully implanted terms, as compared to its prevalence across all ranges of implantation[54].

[54] Due to the low number of fully implanted terms, no graph is presented in this sub-section.

4.5. Distribution by pattern and implantation

In order to provide a synthesis of data presented in the previous sections that can allow an examination of the possibility of correlations between term formation patterns and term implantation, this section presents a juxtaposition of data relating to the distribution of terms based on term formation patterns and implantation rates. To this effect, Figure 18 shows the distribution of term formation patterns overall as well as by implantation, on the basis of four ranges of implantation: 1) no implantation, 2) lower implantation, 3) higher implantation and 4) fully implanted terms. The percentages indicate the frequency of the pattern in each range.

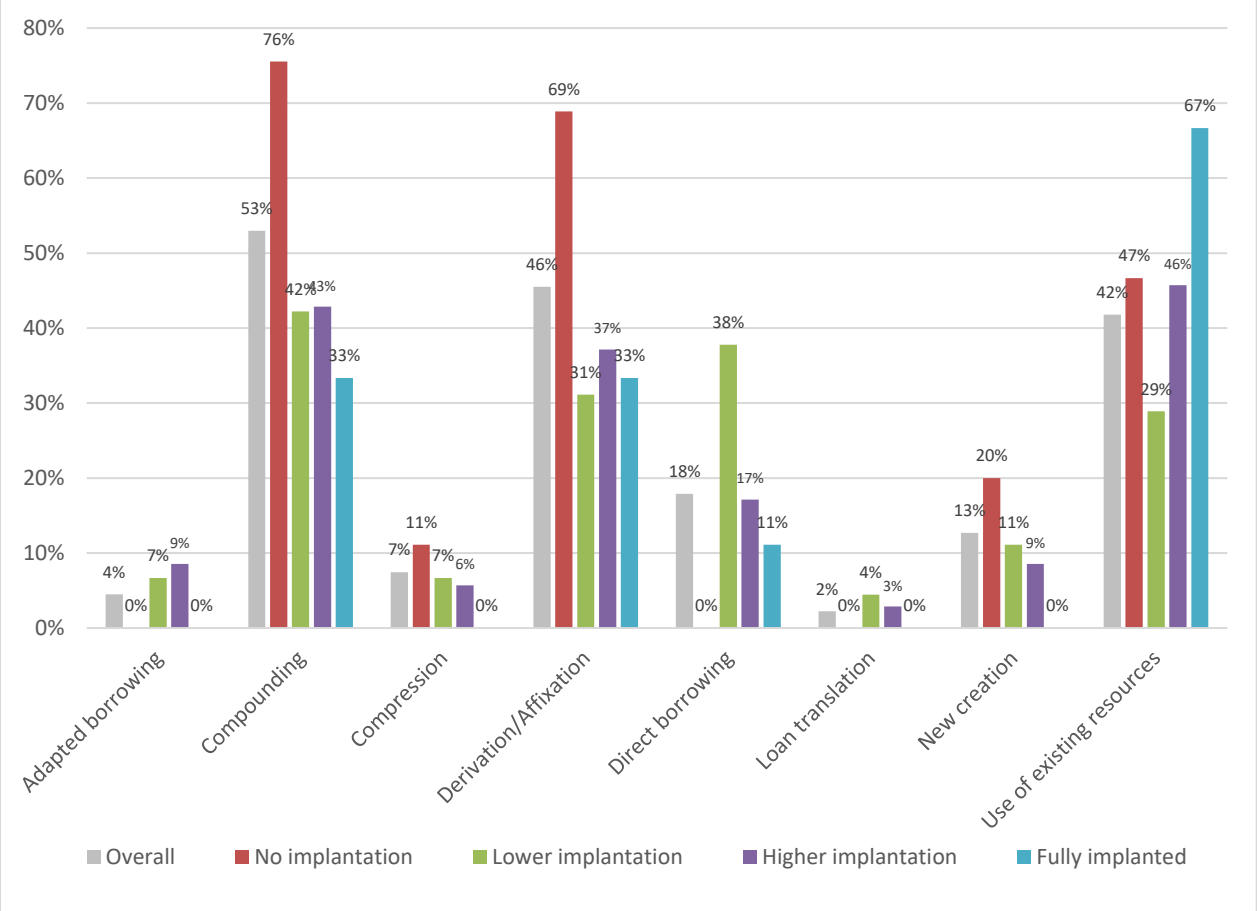

Figure 18 - Frequency of term formation patterns overall and in four ranges of implantation

This graph can be useful for comparing term formation patterns on the basis of their frequencies in each range. It can also provide insight into the frequency of individual patterns across ranges of implantation. For example, if any pattern is more frequently observed in higher ranges of implantation compared to its frequency overall or in lower ranges, this can be interpreted as a favourable factor for the pattern. Likewise, if any pattern is more frequently observed in lower ranges of implantation compared to its frequency overall or in higher ranges, this can be interpreted as a negative factor for the pattern.

What the graph indicates, first of all, is that four out of eight term formation patterns are most frequently observed among terms with no implantation. The highest frequencies among terms with no implantation are observed for the compounding (76%) and derivation/affixation (69%) patterns, which also exhibit the greatest difference in frequency compared to the overall frequencies of the patterns: both patterns are 23% more frequently observed among terms with no implantation compared to their overall frequency.

Three of the remaining four patterns are patterns of borrowing (adapted borrowing, direct borrowing, loan translation), which exclude terms with no implantation as previously explained. The only term formation pattern that is not a pattern of borrowing and that has a lower frequency among terms with no implantation compared to its frequency in other ranges of implantation is the use of existing resources. The latter pattern stands out as well with the highest rates of occurrence among terms with higher implantation and fully implanted terms. Finally, the use of existing resources is the only pattern that is more frequently observed among terms with higher implantation and fully implanted terms compared to its overall frequency.

In the lower implantation range, one column that stands out represents direct borrowings with lower implantation, which has the highest frequency observed for this pattern among all

ranges of implantation. Direct borrowings are observed 20% more frequently among terms with lower implantation compared to their frequency overall. While terms exhibiting this pattern constitute 18% of terms studied (24 out of 134 terms), they make up 38% of terms with lower implantation (17 out of 45 terms). This is the greatest difference observed between the two respective frequency rates across all term formation patterns.

In order to complement the examination presented above, which is based mainly on frequency values, Figure 19 (below) illustrates the scatter distribution of all terms, indicating term formation patterns and IC values. This information can be a useful addition to the examination of frequencies, especially for term formation patterns and ranges of implantation that include a relatively lower number of terms.

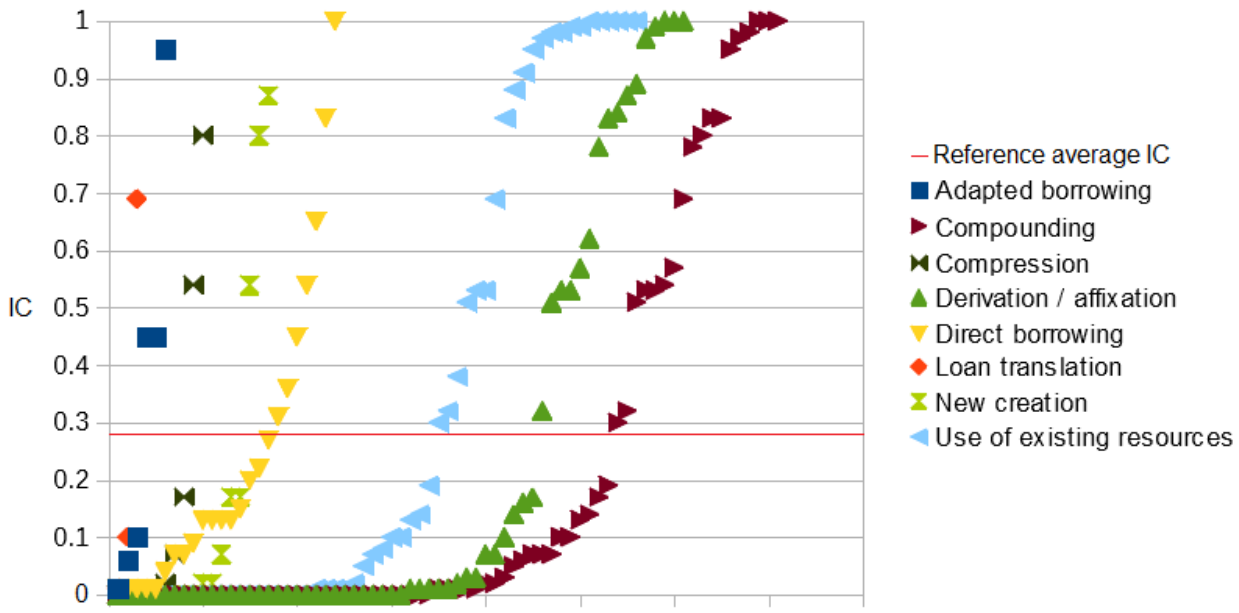

Figure 19 - Scatter distribution of terms by term formation pattern and implantation

The vertical axis represents the level of implantation on the basis of IC values. Each shape and colour combination represents a term formation pattern as indicated in the legend, and each mark

on the chart represents a term exhibiting one of the term formation patterns. It is worth remembering here that term formation patterns are not mutually exclusive, i.e. one term can exhibit more than one term formation pattern and can therefore appear more than once. Note that the horizontal axis does not indicate any value and is only used to spread out the representation of terms and formation patterns for facilitating the interpretation of the figure.

Most of the concentration across all term formation patterns happens in the very low ranges of implantation. This coincides with the high number of terms with no or lower implantation and the subsequent low average IC values. Several term formation patterns can be seen with sporadic occurrences across the vertical axis. Patterns of borrowing (direct borrowing, adapted borrowing, loan translation) are a good example of this, despite having the three highest average IC values (as demonstrated previously in Table 3, p. 72). The relative concentration of direct borrowings in the lower implantation range echoes the observations from the previous figure (Figure 18).

Three term formation patterns appear with some concentration in higher levels of the vertical axis: use of existing resources, compounding and derivation/affixation. This coincides with the frequencies that are outlined in Figure 18 above (p. 95), as these three patterns are the most frequently observed patterns among terms with higher implantation and fully implanted terms with large margins compared to the frequencies observed for other patterns in these higher ranges of implantation.

5. DISCUSSION OF FINDINGS

This chapter provides an interpretation and discussion of the findings presented in the previous chapter. In line with the research question, hypothesis and objectives of this study, the discussion focuses on assessing correlations between term implantation and term formation patterns. Section 5.1 provides a summary of the findings as they directly relate to the research question, and it outlines the correlations observed between term implantation and term formation patterns. Section 5.2 presents emergent findings that can provide a basis for further exploration in future studies.

5.1. Correlations between term implantation and term formation patterns

The data presented in the previous chapter reveals a number of correlations between term implantation and term formation patterns. It should be noted that such correlations do not necessarily imply causation. In other words, any tendency in term implantation that may be observed for a term formation pattern or a group of term formation patterns may be influenced by several other factors as indicated under the scope and limitations (p. 9). Nevertheless, the findings of this study indicate correlations that may be worth exploring further by trials in localization settings or future research that can assess additional factors that influence term implantation.

It is worth acknowledging that, overall, the analysis did not reveal any direct correlations that can be associated with a particular term formation pattern. Rather, the observations apply to two groups of term formation patterns that correlate with two ranges of implantation, as explained below. Some indications that relate to two term formation patterns (namely, compression and borrowing) are also noted.

Patterns with higher and lower implantation

The analysis of data juxtaposing term formation patterns and IC values (as illustrated in Figure 18 and Figure 19, on p. 95 and p. 97 respectively) reveals two groups of term formation patterns. On one hand, patterns of compounding, derivation/affixation, and the use of existing resources exhibit higher frequencies overall and constitute larger portions of terms with higher implantation and fully implanted terms. On the other hand, new creations, compressed terms and borrowings (including direct borrowings, adapted borrowings and loan translations) exhibit lower frequencies overall and they have few or no occurrences among terms with higher implantation and fully implanted terms. In other words, terms exhibiting the patterns in the first group are more common and have higher implantation rates in the corpus analyzed in this study, whereas the patterns in the second group are not as common in term formation and terms exhibiting these patterns have relatively lower implantation rates in the corpus.

The observation of compounding and derivation/affixation in the first group aligns with Sager's remark on patterns of modification of existing resources constituting the most common method in term formation (Sager 1990: 72). Through the use of terminometric research methodology, the findings of this study demonstrate that patterns of modification of existing resources (with the exception of compression) and the use of existing resources indeed appear more frequently, and with higher implantation. The patterns observed in the second group (with the exception of compression) with lower rates of occurrence and lower implantation correspond to the category of patterns Sager collects under "creation of new lexical entities" (ibid. 79).

The main difference identified between the two groups relates essentially to whether existing linguistic resources are used in term formation or new lexical items are introduced to the language to designate new concepts. From a research perspective, this warrants further

investigation. For instance, the corpus compiled for the purposes of this study includes sufficient information about posting dates that would allow a diachronic analysis of term implantation to assess potential changes in implantation of terms over time within the six-year period covered in the corpus. Alternatively, a comparable corpus compiled at a future date could provide evidence of changes in implantation over time and be used to compare the findings of this study with the analysis of a different time period. This type of research could verify whether the implantation of terms exhibiting patterns of creation of new lexical entities improves over time relative to the implantation of terms formed through patterns of modification of existing resources or the use of existing resources.

From a practical perspective, the findings relating to the two groups of term formation patterns can be used in localization settings in an attempt to favour implantation by prioritizing the use of terms exhibiting patterns with higher frequency and implantation, and/or by avoiding the use of terms exhibiting patterns with lower frequency and implantation to some extent. Evaluations of the outcome of such attempts could be carried out to assess whether the implantation trends of term formation patterns as identified in this study are reflected in users' reactions in targeted terminological trials or focus groups for specific software products, for example.

Compressed terms

As the only pattern of modification of existing resources with a lower frequency of occurrence and lower implantation rates, the compression pattern presents an interesting picture when observed more closely. Only two out of 10 compressed terms appear among terms with higher implantation. Meanwhile, some compressed English terms that are directly borrowed in French were observed, in the course of the analysis stage, to have higher implantation rates. Within the

scope of this study, these terms are assessed as direct borrowings, since the term formation pattern that applies to their use in French is, indeed, direct borrowing. However, the higher implantation levels of compressed English terms directly borrowed in French, in relation to compressed French terms, is worthy of closer examination in order to compare the prevalence of this pattern in these two languages. Table 5 below lists these borrowed terms and their IC values.

Term	**IC**
blog	0.54
OS	0.45
CPU	0.36
e-mail	0.15
syslog	0.13
ISP	0.07

Table 5 - Compressed English terms directly borrowed in French

Among the 24 direct borrowings observed in the corpus, the six terms listed in the table above are compressed English terms, and three of them are terms with higher implantation. Notwithstanding the small sample sizes, compressed French terms with higher implantation are observed in two of 10 cases (20%) and compressed English terms directly borrowed in French with higher implantation are observed in three out of six cases (50%). This difference indicates that the lower rates of occurrence and implantation observed for compressed terms in French may, aside from other potential factors, be linked to the fact that this pattern does not seem to be as common in French as it may be in English. For instance, Picone (1992: 19) lists "l'exploitation de la siglaison" as one of the newer methods of lexical creation observed in French as a result of the increased influence of English as a global language and a linguistic model. Moreover,

Wolosin's (1996) corpus-based study of French specialized dictionaries and publications on the information technology domain reveals that 96% of the acronyms and abbreviations in the corpus are in English. The findings of this study regarding compressed terms complemented by existing literature on the use of compressed terms in French carry clear indications of a lower likelihood that terms exhibiting this pattern would reach high levels of implantation. Although this is a language-specific observation, it can inform localization decisions for improved implantation, terminology standardization, and eventually improved reception of localized content in French.

Borrowed terms

The lower frequencies and rates of implantation observed for patterns of borrowing can help form the basis of an argument against the use of these patterns, since they have long created adverse reactions in linguistic communities (Humbley 1997: 263). Reaction to linguistic interference may somewhat habitually stem from nostalgia or a yearning for linguistic purity (Laforest 2002: 82). However, the data provided in this study indicates on the basis of scientific methods that the borrowed terms examined perform poorly with regard to term implantation. Due to the small sample size, it is not possible to reach a wide-reaching conclusion. Nevertheless, this assessment can inform terminological decisions made by members of any linguistic community who are motivated to contribute to the evolution of their language and its capacity to designate new concepts.

5.2. Emergent findings

This section outlines the emergent findings as they relate to the theoretical framework of the thesis. These findings stem from the examination of the distribution of the terms and concepts

based on the source in which they are observed, as this carries indications regarding the use (in other words, implantation) or storage of terminology in each source.

A first emergent finding is that the terms and concepts found in the corpus carry indications regarding term implantation in discussions of open-source software within the Ubuntu-Québec community. Table 6 outlines the terms per concept observed in the corpus sample and in the terminological databases consulted in this study.

<table>
<tr><td colspan="2">252 terms in the sample, designating 180 concepts
(1.4 terms per concept)</td></tr>
<tr><td>55 terms from the sample,
designating 37 concepts
(1.5 terms per concept)</td><td>79 terms from databases,
designating 37 concepts
(2.1 terms per concept)</td></tr>
<tr><td colspan="2">134 terms designating 37 concepts
(3.6 terms per concept)</td></tr>
</table>

Table 6 - Terms per concept observed in the corpus and in the terminological databases

The terminometric analysis of the 134 terms originating in the corpus sample and in three terminological databases reveal that there are approximately 3.6 terms on average for each one of the 37 concepts examined. However, the average number of competing terms is lower within the corpus sample (as explained in sub-section 3.2.3): 252 terms identified in the corpus sample designate a total of 180 concepts. This means 1.4 terms per concept on average. Of the 252 terms initially identified from the sample, 55 figure among the 134 terms that are ultimately selected for further analysis, designating 37 concepts. Thus, the average number of terms per concept within this sub-group is approximately 1.5. The average IC within this sub-group is also noticeably higher than the reference average IC of 0.28, at 0.58. This can partly be explained by the fact that the presence of these terms in the sample naturally guarantees their presence in the

corpus. Finally, the average numbers of terms per concept and the average IC values cited above for terms observed in the corpus generally compare favourably to the results obtained in several terminometric studies to date (cf. Karabacak 2009; Quirion 2003b; Quirion 2011).

The coverage of terms and concepts stored in the three terminological databases consulted in this study reflects the differences between the coverage and intended purposes of these resources. For instance, greater terminological variety is observed in TERMIUM and the GDT. In other words, these resources cover a greater number of terms per concept on average. This does not necessarily come as a surprise since both of these resources are large-scale terminology databases that strive to provide exhaustive terminological information for the general public and specialized users alike. On the other hand, the terminology used in the corpus is more closely reflected in the MS Language Portal, which provides standardized, concise and consistent terminological information that represents the terminology used in Microsoft products. Thus, the latter resource seems more practically relevant in this context, although the value of exhaustive terminological coverage, which comes with the more elaborate supporting information stored in term records, cannot be disregarded.

Finally, the effective coverage of terminology used in the discussion of open-source software in a terminological resource maintained by Microsoft, i.e. a commercial software brand, is indicative of the influence of mainstream commercial software on the terminology of the domain. It is worth remembering here that the MS Language Portal also enables users to provide input through terminology forums. The implantation of terms within the discussions of the Ubuntu community and the implantation of terms found in the MS Language Portal are strong indications supporting Quirion's observation regarding the favourable effect that user participation in terminological decision-making has on term implantation (Quirion 2004: 197).

6. CONCLUSION AND FINAL REMARKS

As presented in the Introduction and Literature Review chapters (1 and 2), the theoretical framework of this thesis is built upon three fields: terminology, localization and user-generated content. The terminological framework of Sager (1990) is used as a means of identifying term formation patterns in order to sort terms for the purposes of terminometric measurement, thus applying a particular aspect of terminology theory in a new setting. The terminometric research methodology proposed by Quirion (2003a) is adapted for use in a new context as well, i.e. for a study that is conducted outside the context of language planning, shifting the focus from institution to community. This is a necessary shift in focus considering the ongoing global shift in content production and distribution, which is increasingly influenced by online communities. Finally, the examination of users' discussions related to localized open-source software through a terminological lens is motivated by the growing role assumed by users, and it addresses the subsequent need to understand the numerous aspects of open collaboration.

6.1. Limitations

The scope and limitations of this study apply to the analysis of the data collected for this project, the discussion based on this analysis, and the findings that emerge. Below is a summary of the findings of the study in light of these constraints.

Due to the nature and pace of technological development, observations on the concepts discussed in online forums, on users' terminological preferences in these discussions, and on the relevance of Ubuntu as an open-source desktop operating system risk seeming outdated. However, in light of the literature review presented in this document, there is reason to believe that examining language, as it is used in an authentic context that represents a growing field of interest and activity, can reveal useful information for research and can potentially inform

practices within the field.

The findings related to correlations between term formation patterns and term implantation are limited to the corpus and the terms examined in this study. The identification of term formation patterns is affected by the limitations of the identification process, which relies on how the definitions of each pattern are presented in this document and how they are interpreted and applied by each of the three individuals (including the author) carrying out the identification process. For future studies, in order to enhance the reliability and consistency of the outcomes, this approach could be improved by establishing detailed guidelines for identifying each pattern and by obtaining assistance from individuals who have a higher level of familiarity with or expertise in terminology. The terms examined in this study reflect the terminology used in the discussion that takes place within the Ubuntu-Québec community and not necessarily the terminology used in the software interface or peripheral material such as help documents. However, considering the important role played by discussion forums in the context of open-source software, such discussion is likely to be representative of relevant software terminology.

One of the findings regarding a specific term formation pattern concerns borrowings and their lower rates of frequency and implantation, which may justify avoiding the pattern in favour of improved implantation. It is worth remembering here that only English terms that were found in the corpus were examined under borrowings, and that English terms not found in the corpus were excluded from the study. Alternatively, if non-French terms stored in TERMIUM and the GDT under French term records (for example, those that are marked with the usage label "AVOID") without occurrences in the corpus had not been excluded from the study (as explained in sub-section 3.2.3), the analysis could have generated a more informative picture of the use and

implantation of borrowed terms. This could have eliminated or at least reduced the instances of average IC values or distribution frequency rates; at times, this seemed positive, but that conclusion could be misleading. The analysis and interpretation of this type of data was carried out with careful consideration of this limitation.

Finally, in contrast to an intuition that the discussion forums would include topics relating specifically to localization or terminology, the corpus did not include any examples of user input directly addressing these aspects of open-source software localization. Considering the vast amount of communication recorded on discussion forums, it is likely that such discussion does take place on discussion forums but that – for a variety of possible reasons, related to the time period covered in the corpus, to the community selected for the study, etc. – it was not included in the corpus. Examining discussion of such topics could also help shed more light on decisions made throughout the localization process, which would lead to a better understanding of open-source software localization and its terminological outcome.

6.2. Contribution to knowledge, reflections for future research

The findings of this thesis highlight the need for further diachronic study, but they also contribute to understanding un(der)explored factors that may influence term implantation. There is sufficient evidence to suggest that term formation patterns may indeed be a factor in term implantation. The findings suggest that terms formed through patterns of modification of existing resources (with the exception of compressed terms) and through the use of existing resources are more common and have higher rates of implantation compared to terms formed through patterns of creation of new lexical entities such as borrowing or neologism. The implantation of new lexical entities that are formed through these patterns in the context of localized software may change or improve over time, relative to terms formed through patterns of modification of

existing resources and the use of existing resources; however, this remains to be tested.

The examination of the sources of the terms and concepts analyzed in this study reflect the differences in the coverage of terminology in the terminological resources that were consulted. Terminological databases created by public institutions for use by language professionals and the general public (TERMIUM, GDT) cover a vastly greater number of terms, while the terminological database of a mainstream software producer (the MS Language Portal) reflects a notable amount of terminological standardization. Moreover, there is a high level of correspondence between the terms used in the corpus and the terms stored in the MS Language Portal. This indicates that user participation in terminological decision-making creates a positive outcome regarding term implantation, and that mainstream software terminology influences the terminology of the domain.

By creating a theoretical framework and a methodology based on existing research in terminology, localization, and user-generated content, this project paves the way for future research within the intersection of these fields. This new avenue presents significant potential because it relates to subjects that promise to occupy an increasingly large portion of research on knowledge and technology production and distribution worldwide, as the scope of localization continues to expand, as passive users increasingly become active contributors, and as locales increasingly gain access to spreading technology. The methodology used in this project could potentially influence future studies, which could in turn help gradually fill the gap in localization research that focuses on the reception of localized products and the outcome of user-generated translation. In addition, as this project applies terminometric research methods to a new context outside of language planning, it can help pave the way for using terminometrics to measure the outcome of terminological choices in a variety of contexts, without being limited to language

planning.

The introduction of an internal reference average IC as a tool to gauge the significance of implantation rates may also prove useful in other terminometric studies as an approach that is founded on the understanding of implantation as relative by nature. This approach facilitates the assessment of the implantation of terms in relation to terms within the same subject field. Thus, it favours an internal point of reference that is objective, rather than arbitrary estimates that are based on subjective opinions or intuitions.

The identification of factors that influence term implantation in the context of localization has the potential to initiate a chain reaction by encouraging decisions that may foster terminology implantation and thus facilitate standardization, consequently leading to improved experiences with localized software. Technology is usually seen as a source of Anglicization, since English is generally the language of distribution, and English terms are commonly borrowed in many languages to express new technical and technological concepts. However, enhanced user experience with localized products can help disseminate new terminology through technology. For languages that depend on localization to import technology, improvements in the reception of localized products can encourage native language use in such fields, thus strengthening linguistic identity. A new perspective on technology, whereby it is seen as a source of linguistic novelty, can increase the capacity of speakers of these languages to define the modern world using their own language, ease their linguistic insecurity and increase the appeal of their mother tongue.

BIBLIOGRAPHY

AUGER, Pierre (1986) "Francisation et terminologie : L'aménagement terminologique", in *Termia 84 : terminologie et coopération internationale : la terminologie, outil indispensable au transfert des technologies*, Rondeau and Sager (eds.), Colloque international de terminologie. Québec : Girsterm. 47-55.

AYDIN, Özgür, Semih Bilgen, Sevgi Çıkrıkçı, İclâl Ergenç, Güneş Müftüoğlu and Melek Yücel (2004) "Yabancı dille yapılan üniversite eğitiminin anadili becerileri üzerindeki etkisi" [Effects of foreign language university education on native language skills], *Bilim ve Ütopya [Science and Utopia],* May 2004.

BAER, Naomi (2010) "Crowdsourcing: Outrage or opportunity?", *Translorial*, February 2010. http://translorial.com/2010/02/01/crowdsourcing-outrage-or-opportunity (last accessed 24 June 2016)

BEATTY, Jeff and Stas Malolepszy (2013) "Open source localization", *MultiLingual*, vol. 24, no. 1, 28-32.

BERNAL MERINO, Miguel (2006) "On the translation of video games", *The Journal of Specialised Translation*, no. 6, 22-36. http://www.jostrans.org/issue06/art_bernal.pdf (last accessed 28 June 2016)

BEY, Youcef, Christian Boitet and Kyo Kaguera (2008) "BEYTrans: A wiki-based environment for helping online volunteer translators" in *Topics in Language Resources for Translation and Localisation*, Rodrigo (ed.), Amsterdam/Philadelphia: John Benjamins. 135-150.

BİLGEN, Barış (2012) "An analysis of localization strategies with the help of terminometrics and translation forums", presentation at the *4th Conference of the International Association for Translation and Intercultural Studies*, Queen's University Belfast, 25 July 2012, Northern Ireland, UK.

BLOODGOOD, Michael and Chris Callison-Burch (2010) "Using Mechanical Turk to build machine translation evaluation sets", Proceedings of *the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, Los Angeles, California, June 2010*. 208-211. http://www.cis.upenn.edu/~ccb/publications/using-mechanical-turk-to-build-machine-translation-evaluation-sets.pdf (last accessed 24 June 2016)

BLUM, Guillaume (2012) "De la liberté du logiciel et de son ouverture : tour d'horizon et perspectives", *Télescope,* vol. 18, no. 1-2, 121-138. http://www.telescope.enap.ca/Telescope/docs/Index/Vol_18_no_1-2/Telv18n1-2_blum.pdf (last

accessed 24 June 2016)

BOUCHARD, Chantal (1989) "Une obsession nationale : l'anglicisme", *Recherches sociographiques*, vol. 30, no. 1, 67-90.

BOURIGAULT, Didier and Monique Slodzian (1999) "Pour une terminologie textuelle", *Terminologies nouvelles*, no. 19, 29-32.

BOWKER, Lynne (2006) "Translation memory and 'text' " in *Lexicography, Terminology and Translation: Text-based Studies in Honour of Ingrid Meyer,* Bowker (ed.)*,* Ottawa: University of Ottawa Press, 175-187.

BOWKER, Lynne and Jennifer Pearson (2002) "Introducing corpora and corpus analysis tools" in *Working with Specialized Language: A Practical Guide to Using Corpora,* London; New York: Routledge, 9-21.

BOWKER, Lynne and Shane Hawkins (2006) "Variation in the organization of medical terms: Exploring some motivations for term choice", *Terminology*, vol. 12, no. 1, 79-110.

CABRÉ CASTELLVI, Maria Teresa (2003) "Theories of terminology. Their description, prescription and explanation", *Terminology*, vol. 9, no. 2, 163-199.

CASTELLS, Manuel (2015) *Networks of Outrage and Hope: Social Movements in the Internet Age,* 2nd edition, Cambridge: Polity, 328 pages.

CORBEIL, Jean-Claude (1980) *L'aménagement linguistique du Québec*, Montréal : Guérin. 154 pages.

CORBOLANTE, Lucia and Ulrike Irmler (1997) "Software Terminology and Localization" in *Handbook of Terminology Management*, Wright and Budin (eds.), Amsterdam/Philadelphia: John Benjamins, vol. 2, 516-535.

DAOUD, Mohammad, Kyo Kageura, Christian Boitet, Asanobu Kitamoto and Daoud Daoud (2010) "Passive and active contribution to multilingual lexical resources through online cultural activities", Proceedings of *the 6th International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE 2010,* Beijing, China, 21-23 August 2010. Hanover: IEEE, 2010.

DAOUST, Denise (1995) "Quelques facteurs sociodémographiques qui sous-tendent le choix des langues pour la terminologie en milieu de travail et leur interrelation avec des attitudes", *Meta*, vol. 40, no. 2, 260-283.

DEPALMA, Donald A. and Nataly Kelly (2011) "Project management for crowdsourced translation: How user-translated content projects work in real life", in *Translation and Localization Project Management: The Art of the Possible,* Dunne, Dunne (eds.), Amsterdam/Philadelphia: John Benjamins, 379-407.

DÉSILETS, Alain (2007) "Translation Wikified: How will massive online collaboration impact the world of translation?", Proceedings of *Translating and the Computer*, *no. 29, 29-30 November 2007, London, UK.* National Research Council of Canada, 2007.

DÉSILETS, Alain and Jaap van der Meer (2011) "Co-creating a repository of best-practices for collaborative translation", *Linguistica Antverpiensia*, vol. 10, 27-46.

DIKI-KIDIRI, Marcel (2007) "Éléments de terminologie culturelle", *Cahiers du Rifal*, no. 26, 14-25.

DISTROWATCH (n.d) Top ten distributions. http://distrowatch.com/dwres.php?resource=major (last accessed 24 June 2016)

DRAY, Susan M. and David A. Siegel (2006) "Melding paradigms: Meeting the needs of international customers through localization and user-centered design" in *Perspectives on Localization*, Dunne (ed.), Amsterdam/Philadelphia: John Benjamins. 281-307.

DUNNE, Keiran J. and Elena S. Dunne (2011) *Translation and Localization Project Management: The Art of the Possible*, Amsterdam/Philadelphia: John Benjamins, 430 pages.

EAGLE, Nathan (2009) "txteagle: Mobile Crowdsourcing" in *Internationalization, Design and Global Development*, Akin (ed.), San Diego: Springer, 447-456.

ESS, Charles and the Association of Internet Researchers ethics working committee (2002) "Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee", *AoIR*. http://www.aoir.org/reports/ethics.pdf (last accessed 16 November 2016)

ESSELINK, Bert (2006) "The evolution of localization" in *Translation Technology and its Teaching*, Pym, Perekstenko, Starink (eds.), Tarragona: Intercultural Studies Group, 21-30.

ESSELINK, Bert (2000) *A Practical Guide to Localization*, Amsterdam/Philadelphia: John Benjamins, 488 pages.

EXTON, Chris, Asanka Wasala, Jim Buckley and Reinhard Schäler (2009) "Micro crowdsourcing: A new model for software localisation", *Localisation Focus*, vol. 8, no. l, 81-89.

FLUIXA, Noemi (2004) "Software Localisation: Outsourcing or in-house?", *Localisation Reader 2004-2005*, Limerick/Sandpoint: LRC, 67-68.

FOLARON, Deborah (2010) "Networking and volunteer translators" in *Handbook of Translation Studies,* Gambier, van Doorslaer (eds.), Amsterdam/Philadelphia: John

Benjamins, 230-234.

FRANK, Leona (2013) "In-house translation teams – still worth the investment?", *MultiLingual,* vol. 24, no. 7, 29-33.

GARCIA, Ignacio and Vivian Stevenson (2009) "Translation trends and the social web", *MultiLingual*, vol. 20, no. 3, 28-31.

GARIÉPY, Julie (2013) *La collaboration en terminographie : étude de cas comparée de la terminographie collaborative et de la terminographie classique.* MA thesis, University of Ottawa.

GAUDIN, François (2007) "Quelques mots sur la socioterminologie", *Cahiers du Rifal*, no. 26, 26-35.

GERBAUDO, Paolo (2012) *Tweets and the streets: social media and contemporary activism*, London: Pluto, 194 pages.

GOUGH, Joanna (2011) "An empirical study of professional translators' attitudes, use and awareness of Web 2.0 technologies, and implications for the adoption of emerging technologies and trends", *Linguistica Antverpiensia*, vol. 10, 195-225.

GUADAGNO, Rosanna E., Daniel M. Rempala, Shannon Murphy, Bradley M. Okdie (2013) "What makes a video go viral? An analysis of emotional contagion and internet memes", *Computers in Human Behavior,* vol. 29, no. 6, 2312-2319.

HAMELIN, Louis-Edmond (1995) "Le québécisme *nordicité*: de la néologie à la lexicalisation." *TTR*, vol. 8, no. 2, 51-65.

HAYTHORNTHWAITE, Caroline (2009) "Crowds and Communities: Light and Heavyweight Models of Peer Production." Proceedings of *the 42nd Hawaii International Conference on System Sciences*, *Los Alamitos, CA*. IEEE Computer Society, 1-10.

HOWE, Jeff (2006) "The Rise of Crowdsourcing", *Wired*, Issue 14.06, June 2006. http://www.wired.com/wired/archive/14.06/crowds.html (last accessed 24 June 2016)

HUMBLEY, John (1997) "Language Planning and Terminology Planning: The Francophone Experience", in *Handbook of Terminology Management,* Wright and Budin (eds.), Amsterdam/Philadelphia: John Benjamins, vol. 1, 261-277.

JIMENEZ-CRESPO, Miguel A. (2013) *Translation and Web Localization,* New York: Routledge, 256 pages.

JIMENEZ-CRESPO, Miguel A. (2011) "From many one: Novel approaches to translation quality in a social network era", *Linguistica Antverpiensia*, vol. 10, 131-152.

KARABACAK, Erkan (2009) "Acceptance of terminology sanctioned by the Turkish Language Society: A study of the use of economic terms in Turkish newspapers", *Terminology*, vol. 15, no. 2, 145-178.

KARSCH, Barbara Inge (2014) "Terminology work and crowdsourcing: Coming to terms with the crowd", in *Handbook of Terminology, Vol.1,* Kockaert, Steurs (eds.), Amsterdam/Philadelphia: John Benjamins, 291-303.

KARSCH, Barbara Inge (2006) "Terminology workflow in the localization process" in *Perspectives on Localization*, Dunne (ed.), Amsterdam/Philadelphia: John Benjamins. 173-191.

KELLY, Nataly, Rebecca Ray and Donald A. DePalma (2011) "From crawling to sprinting: Community translation goes mainstream", *Linguistica Antverpiensia*, vol. 10, 75-94.

KLISCHEWSKI, Rolf (2015) "A case for dedicated games localization tools", *MultiLingual,* vol. 26, no. 4, 44-48.

LAFOREST, Marty (2002) "Attitudes, préjugés et opinions sur la langue" in *Le français, une langue à apprivoiser,* Verreault (ed.), Presses de l'Université Laval, 81-91.

LAKHANI, Karim R. and Eric von Hippel (2002) "How open source software works: 'free' user-to-user assistance", *Research Policy,* vol. 32, 923-943.

LEPOURAS, Giorgos and George R. S. Weir (1999) "It's not Greek to me: Terminology and the second language problem", *SIGCHI Bulletin*, vol. 31, no. 2, 17-24.

LYNCH, Jim (2014) "DistroWatch resolves its domain registrar problems", *ITworld,* 14 July 2014, http://www.itworld.com/article/2832342/distrowatch-resolves-its-domain-registrar-problems.html (last accessed 24 June 2016)

MANGIRON, Carmen and Minako O'Hagan (2006) "Game Localisation: Unleashing Imagination with Restricted Translation", *The Journal of Specialised Translation*, no. 6, 10-21. http://www.jostrans.org/issue06/art_ohagan.pdf (last accessed 24 June 2016)

MARKHAM, Annette and Elizabeth Buchanan (2012) "Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee (version 2.0)", *AoIR,* http://aoir.org/reports/ethics2.pdf (last accessed 16 November 2016)

MAURAIS, Jacques (1993) "Terminology and language planning" in *Terminology: Applications in interdisciplinary communication*, Amsterdam/Philadelphia: John Benjamins, 111-125.

MCDONOUGH DOLMAYA, Julie (2011) "The ethics of crowdsourcing", *Linguistica Antverpiensia*, vol. 10, 97-110.

MESIPUU, Marit (2012) "Translation crowdsourcing and user-translator motivation at Facebook and Skype", *Translation Spaces,* vol.1, 33-53.

MEYER, Christian M. and Iryna Gurevych (2002) "A new rival for expert-build lexicons? Exploring the possibilities of collaborative lexicography" in *Electronic Lexicography,* Granger, Paquot (eds.), Oxford: Oxford University Press, 259-291.

MUNRO, Robert (2010) "Crowdsourced translation for emergency response in Haiti: the global collaboration of local knowledge", Proceedings of *Collaborative Translation: technology, crowdsourcing, and the translator perspective*, *AMTA 2010, 31 October 2010*, *Denver, Colorado*. http://amta2010.amtaweb.org/AMTA/papers/7-01-01-Munro.pdf (last accessed 24 June 2016)

NEEDLEMAN, Rafe (2011) "CloudCrowd: An assembly line for content", *CNET News*, 31 January 2011. http://news.cnet.com/8301-19882_3-20030094-250.html (last accessed 24 June 2016)

NİŞANYAN, Sevan (2009) *Sözlerin Soyağacı. Çağdaş Türkçe'nin Etimolojik Sözlüğü. [Genealogy of Words. An etymological dictionary of modern Turkish.]* Istanbul: Alfa. Online version: www.nisanyansozluk.com, last accessed 24 June 2016.

O'HAGAN, Minako (2011) "Community Translation: Translation as a social activity and its possible consequences in the advent of Web 2.0 and beyond", *Linguistica Antverpiensia*, vol. 10, 11-23.

O'HAGAN, Minako (2009) "Evolution of User-generated Translation: Fansubs, Translation Hacking and Crowdsourcing", *The Journal of Internationalisation and Localisation*, vol. 1, 94-121.

PAVEL, Silvia (1993) "Neology and phraseology as terminology-in-the-making" in *Terminology: Applications in interdisciplinary communication*, Amsterdam/Philadelphia: John Benjamins, 21-34.

PICONE, Michael D. (1992) "Le français face à l'anglais : aspects linguistiques", *Cahiers de l'Association internationale des études françaises,* vol. 44, 9-23. http://www.persee.fr/doc/caief_0571-5865_1992_num_44_1_1775 (last accessed 4 July 2016)

PYM, Anthony (2005) "Localization: On its nature, virtues and dangers" http://usuaris.tinet.cat/apym/on-line/translation/Localization_bergen.doc (last accessed 24 June 2016)

QUIRION, Jean (2011) "Dynamique terminologique et terminométrie", *Terminology,* vol. 17, no.1, 113-133.

QUIRION, Jean and Jacynthe Lanthier (2006) "Intrinsic qualities favouring term implantation: Verifying the axioms", in *Lexicography, Terminology, and Translation: Text-Based Studies in Honour of Ingrid Meyer*, Bowker (ed.), Ottawa: University of Ottawa Press. 107-118.

QUIRION, Jean (2004) "État de la question sur la nature des facteurs d'implantation terminologique", *Colloque international sur la traduction : Traduction et Francophonies. Traduire en Francophonie*, *Rennes (France), Université Rennes 2, 12, 13 septembre 2003*, Paris, Maison du dictionnaire. http://www.colloque.net/archives/2003/volume_1/quirion.pdf (last accessed 24 June 2016)

QUIRION, Jean (2003a) "Methodology for the design of a standard research protocol for measuring terminology usage", *Terminology*, vol. 9, no. 1, 29-49.

QUIRION, Jean (2003b) *La mesure de l'implantation terminologique : proposition d'un protocole. Étude terminométrique du domaine des transports au Québec*, coll. Langues et sociétés, no. 40, Montréal : Office québécois de la langue française. 225 pages.

RAYMOND, Eric (1999) *The Cathedral and the Bazaar*, O'Reilly: Cambridge. 269 pages.

ROUSSEAU, Louis-Jean (1993) "Terminology and languages in contact in Québec" in *Terminology: Applications in interdisciplinary communication*, Amsterdam/Philadelphia: John Benjamins, 35-41.

SAGER, Juan C. (1997) "Term formation" in *Handbook of Terminology Management*, Wright and Budin (eds.), Amsterdam/Philadelphia: John Benjamins, vol. 1, 25-41.

SAGER, Juan C. (1990) *A Practical Course in Terminology Processing*, Amsterdam/Philadelphia: John Benjamins, 266 pages.

SCHÄLER, Reinhard (2008) "Linguistic resources and localization" in *Topics in Language Resources for Translation and Localisation*, Rodrigo (ed.), Amsterdam/Philadelphia: John Benjamins. 195-214.

SCHMITZ, Klaus-Dirk (2014) "Terminology and localisation" in *Handbook of Terminology, Vol.1,* Kockaert, Steurs (eds.), Amsterdam/Philadelphia: John Benjamins, 451-463.

SHIRKY, Clay (2012) "Clay Shirky: How the Internet will (one day) transform government", *TEDGlobal 2012,* TED Talks.

http://www.ted.com/talks/clay_shirky_how_the_internet_will_one_day_transform_government, filmed June 2012, last accessed 24 June 2016.

SHIRKY, Clay (2008) "Clay Shirky: Institutions vs. Collaboration", *TEDGlobal 2005*, TED

Talks, http://www.ted.com/talks/clay_shirky_on_institutions_versus_collaboration.html, filmed July 2005, posted July 2008, last accessed 24 June 2016.

SOUPHAVANH, Anousak and Theppitak Karoonboonyanan (2005) *Free/Open Source Software: Localization.* New Delhi: Elsevier & UNDP-APDIP, 58 pages. http://www.iosn.net/l10n/foss-localization-primer/foss-localization-primer.pdf (last accessed 8 July 2015); Wikibook version available at: https://en.wikibooks.org/wiki/FOSS_Localization (last accessed 24 June 2016)

STATCOUNTER (2016) Top 7 Desktop, Tablet & Console OSs from 2011 to 2016. http://gs.statcounter.com/#os-ww-yearly-2011-2016 (last accessed 8 August 2016)

STATCOUNTER (2013) Top 7 Operating Systems from 2011 to 2013. http://gs.statcounter.com/#os-ww-yearly-2011-2013 (last accessed 24 June 2016)

SUROWIECKI, James (2004) *The Wisdom of Crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*, New York: Doubleday. 296 pages.

TEMMERMAN, Rita (2000) "Une théorie réaliste de la terminologie : le sociocognitivisme", *Terminologies nouvelles*, no. 21, 59-65.

VAN GENABITH, Joseph (2009) "Next generation localisation", *Localisation Focus*, vol. 8, no. 1, 4-10.

VAUGHAN-NICHOLS, Steven J. (2014) "Linux's DistroWatch site stumbles", *ZDNet – Networking,* 6 July 2014, http://www.zdnet.com/article/linuxs-distrowatch-site-stumbles/ (last accessed 24 June 2016)

WARRELL, Jacqueline G. and Michele Jacobsen (2014) "Internet research ethics and the policy gap for ethical practice in online research settings", *Canadian Journal of Higher Education,* vol. 44, no.1, 22-37.

WOLOSIN, Claudia (1996) "Problèmes de traduction posés par la siglaison dans le domaine des nouvelles technologies de l'information et de la communication", *ASp,* vol. 11-14, 147-160. http://asp.revues.org/3468 (last accessed 4 July 2016)

ZIMMER, Michael (2010) "'But the data is already public': on the ethics of research in Facebook", *Ethics and Information Technology,* vol.12, 313-325.

APPENDIX 1: Definitions

The definitions are presented based on a conceptual hierarchy under three main groups corresponding to the three fields of literature that form the theoretical framework of this study: localization, user-generated content, terminology.

Localization

- *Globalization:* In the context of software localization, globalization refers to all activities involved in rendering a software product global, as in launching multiple localized versions across the globe (O'Hagan 2005: 1)
- *Internationalization*: Internationalization refers to measures taken in software development to provide a level of linguistic and cultural neutrality to the fundamental structure of the software in order to allow for the modifications required for localization, and to avoid subsequent modifications on the original program for any specific locale (Esselink 2000:2; Schmitz 2014: 451).
- *Localization*: Esselink defines localization as "the translation and adaptation of a software or web product, which includes the software application itself and all related product documentation," for a specific target locale, representing "a specific combination of language, region, and character encoding" (Esselink 2000: 1).
- *Translation*: Translation is frequently mentioned in localization literature as one of the activities that takes places within the process of localization (cf. Esselink 2006: 24; Bernal Merino 2006: 29). Within the limits of this project, the term *translation* is understood as designating one of the many components of localization, and is used to refer specifically to the translation of text, and only text, as opposed to *localization* which encompasses textual as well as extra-textual, multimedia elements.

User-generated content

- *Collaborative:* Within the context of this study, the term *collaborative* is used in relation to worldwide human collaboration taking place on the Internet, as discussed by Désilets (2007). Accordingly, the term is used to designate platforms or initiatives involving the participation of multiple individuals in a web-based setting with a shared goal, which can range from sharing content to undertaking specific tasks in a shared manner.
- *Content*: Within the context of this project, the term *content* is used to refer to any type of material or service published or provided online, including but not limited to image, text, video, audio and software.
- *Contributor*: The term *contributor* is used to refer to users who contribute to a collaborative initiative in any way.
- *Open*: The term *open* is used in relation to its sense of peer production as discussed by Haythornthwaite (2009), and refers to unrestricted access by users of a software product or website to facilities that enable the modification of its various components. The term *open source (open-source,* when used as adjective) is used to refer specifically to a software product which has a source code that enables modification.
- *User*: In the context of this project, the term *user* refers to an individual who uses computers and computer applications (including websites and all types of software).

Terminology

- *Term:* Sager (1990: 57) defines a *term* as a linguistic representation of a concept in specialized language which is based on systematized rules or principles of designation.
    - *Standardization:* In the context of terminology, *standardization* refers to the ideal

of establishing a single term designating each specific concept, essentially based on the traditional theory of terminology which aims "to eliminate ambiguity from technical languages (…) in order to make them efficient tools of communication" (Cabré Castellvi 2003:165).

- *Term formation:* Term formation is understood as the process according to which a concept in specialized language is named, based on systematized rules or principles of designation (Sager 1990: 57).
  - *Pattern*: The term *pattern* is used in reference to Sager's (1990) patterns of term formation. (See sub-section 2.3.2 and section 3.1 for details.)

- *Terminometric research:* Quirion (2003a: 30) defines terminometric research as research that is "designed to measure the degree of implantation of all designations referring to a single concept or to a set of concepts."
  - *Competing term:* This refers to one of the terms designating a notion that is not designated by a single, standardized term. This applies to most terms in the context of this study due to the predominance of non-standardized terminology in the subject field in question (cf. Sager 1990: 82; Lepouras and Weir 1999). The term *competing* is deliberately used instead of *synonymous* since it allows for the inclusion of terms in different languages (English and French terms in this study) within the same language's discourse, and carries the implication of non-standardization.
  - *Implantation*: According to Auger (1986: 51), the *implantation* of terminology refers to the adoption of terms by language users to express specialized concepts, by using them actively in authentic communicative instances.

- *Implantation coefficient (IC)*: Based on Quirion's (2003a) formula, the *implantation coefficient* refers to the relative frequency of a term used to designate a concept in relation to all other competing terms that designate the same concept. (See sub-section 3.2.5 for details.)

APPENDIX 2: Concepts and terms gathered for querying the corpus

Overarching concept	Term	Language	Source(s)
blog	blog	En	GDT, MS, Sample, TERMIUM
blog	blogue	Fr	GDT, Sample, TERMIUM
blog	carnet web	Fr	GDT, TERMIUM
blog	cybercarnet	Fr	GDT, TERMIUM
blog	weblog	En	TERMIUM
button	bouton	Fr	GDT, MS, Sample, TERMIUM
button	bouton de fonction	Fr	TERMIUM
button	bouton poussoir, bouton-poussoir	Fr	TERMIUM
button	button	En	GDT, TERMIUM
button	pushbutton	En	TERMIUM
button	touche de fonction	Fr	TERMIUM
command	command	En	GDT, TERMIUM
command	commande	Fr	GDT, MS, Sample, TERMIUM
command	instruction	Fr	TERMIUM
cursor	curseur	Fr	GDT, MS, Sample, TERMIUM
cursor	cursor	En	GDT, MS, TERMIUM
cursor	marqueur	Fr	TERMIUM
default	by default	En	GDT, TERMIUM
default	default	En	TERMIUM
default	implicite	Fr	TERMIUM
default	par défaut	Fr	GDT, MS, Sample, TERMIUM
disk space	disk space, diskspace, disc space	En	GDT, TERMIUM
disk space	espace disque	Fr	GDT, MS, Sample, TERMIUM
display	affichage	Fr	GDT, Sample, TERMIUM
display	affichage numérique	Fr	GDT
display	display	En	GDT, MS, TERMIUM
display	visualisation	Fr	GDT, TERMIUM
download	download	En	GDT, MS, Sample,

			TERMIUM
download	télécharger, téléchargement	Fr	GDT, MS, Sample, TERMIUM
e-mail	CÉ	Fr	TERMIUM
e-mail	courriel	Fr	GDT, MS, Sample, TERMIUM
e-mail	courriél	Fr	GDT
e-mail	courrier électronique	Fr	GDT, MS, TERMIUM
e-mail	electronic mail	En	TERMIUM
e-mail	electronic message	En	GDT
e-mail	e-mail, email	En	GDT, MS, Sample, TERMIUM
e-mail	e-message	En	GDT
e-mail	mail	En	GDT, Sample
e-mail	mél	Fr	GDT
e-mail	message électronique	Fr	GDT, TERMIUM
Internet provider	access provider	En	GDT
Internet provider	FAI	Fr	GDT, TERMIUM
Internet provider	fournisseur d'accès	Fr	GDT
Internet provider	fournisseur d'accès Internet, fournisseur d'accès à Internet	Fr	GDT, MS, TERMIUM
Internet provider	fournisseur de services Internet	Fr	GDT, MS, TERMIUM
Internet provider	fournisseur d'Internet	Fr	Sample
Internet provider	fournisseur Internet	Fr	GDT, MS, TERMIUM
Internet provider	FSI	Fr	GDT, TERMIUM
Internet provider	IAP	En	GDT
Internet provider	Internet access provider	En	GDT, TERMIUM
Internet provider	Internet provider	En	GDT
Internet provider	Internet service provider	En	GDT, TERMIUM
Internet provider	ISP	En	GDT, TERMIUM
Internet	prestataire d'accès Internet,	Fr	TERMIUM

provider	prestataire d'accès à Internet		
Internet provider	prestataire de service	Fr	GDT
Internet provider	prestataire de services Internet	Fr	GDT
Internet provider	prestataire Internet	Fr	TERMIUM
Internet provider	provider	En	GDT
Internet provider	service provider	En	GDT
interoperability	interfonctionnement	Fr	GDT, TERMIUM
interoperability	interopérabilité	Fr	GDT, MS, Sample, TERMIUM
interoperability	interoperability	En	GDT, TERMIUM
keyboard	clavier	Fr	GDT, MS, Sample, TERMIUM
keyboard	keyboard	En	GDT, TERMIUM
keyboard	keyset	En	TERMIUM
light	léger, légère	Fr	MS, Sample
light	light	En	MS, Sample
log	console listing	En	TERMIUM
log	journal, journal système, journal de marche, journal d'exploitation	Fr	Sample, TERMIUM
log	log	En	Sample
log	registre	Fr	MS
log	syslog	En	TERMIUM
login	connexion, se connecter	Fr	TERMIUM
login	entrée dans le système	Fr	TERMIUM
login	login, log in, log-in	En	GDT, Sample, TERMIUM
login	ouverture de session, ouvrir une session	Fr	GDT, Sample, TERMIUM
login	sign in, sign-in	En	GDT, TERMIUM
memory	dispositif de stockage	Fr	TERMIUM
memory	mémoire	Fr	GDT, MS, Sample, TERMIUM
memory	memory	En	GDT, MS, TERMIUM
memory	storage	En	GDT, TERMIUM
memory	storage device	En	GDT, TERMIUM
memory	store	En	TERMIUM

memory	unité de stockage	Fr	TERMIUM
microprocessor	CPU	En	Sample
microprocessor	microprocesseur	Fr	GDT, MS, Sample, TERMIUM
microprocessor	microprocessing unit	En	TERMIUM
microprocessor	microprocessor	En	TERMIUM
microprocessor	MPU	En	TERMIUM
microprocessor	processeur	Fr	GDT, Sample
motherboard	carte de connexion	Fr	TERMIUM
motherboard	carte maîtresse	Fr	Sample, TERMIUM
motherboard	carte principale	Fr	TERMIUM
motherboard	carte mère, carte-mère	Fr	GDT, MS, TERMIUM
motherboard	main board, main circuit board	En	TERMIUM
motherboard	mobo	En	GDT
motherboard	motherboard	En	GDT, TERMIUM
motherboard	mother card, mothercard	En	TERMIUM
multiboot	amorçage multiple	Fr	GDT
multiboot	démarrage multiple	Fr	MS
multiboot	multiboot, multi-boot	En	Sample
office suite	environnement bureautique	Fr	Sample
office suite	environnement de bureau	Fr	Sample
office suite	OA suite	En	TERMIUM
office suite	office application suite	En	GDT
office suite	office automation suite	En	TERMIUM
office suite	office productivity suite	En	TERMIUM
office suite	office suite	En	GDT, TERMIUM
office suite	productivity suite	En	GDT
office suite	suite bureautique, suite de bureautique	Fr	GDT, TERMIUM
office suite	suite office	Fr	Sample
office suite	suite Office	Fr	MS
online	connecté, en connecté	Fr	GDT, TERMIUM
online	en direct	Fr	TERMIUM
online	en ligne	Fr	MS, Sample, TERMIUM
online	online, on-line	En	TERMIUM
operating system	command system	En	TERMIUM

operating system	executive system	En	TERMIUM
operating system	logiciel d'exploitation	Fr	GDT
operating system	operating system	En	GDT, TERMIUM
operating system	OS	En	MS, Sample, TERMIUM
operating system	SE	Fr	GDT
operating system	système d'exploitation	Fr	GDT, MS, Sample, TERMIUM
panel	écran utilitaire	Fr	TERMIUM
panel	panel	En	TERMIUM
panel	panneau	Fr	MS, Sample
partition (drive)	division	Fr	TERMIUM
partition (drive)	partition	Fr	GDT, MS, Sample, TERMIUM
partition (drive)	partitionnement	Fr	Sample, TERMIUM
programming	codage	Fr	GDT
programming	coding	En	GDT
programming	programmation	Fr	GDT, MS, Sample, TERMIUM
programming	programming	En	GDT, TERMIUM
remote desktop	bureau à distance	Fr	MS, Sample
remote desktop	bureau distant	Fr	Sample
remote desktop	remote desktop	En	MS
removable	amovible	Fr	MS, Sample
removable	démontable	Fr	GDT, TERMIUM
removable	demountable	En	GDT, TERMIUM
removable	detachable	En	GDT, TERMIUM
removable	détachable	Fr	GDT, TERMIUM
removable	removable	En	GDT, TERMIUM
removable	supprimable	Fr	TERMIUM
script	scenario	En	TERMIUM
script	scénario	Fr	TERMIUM
script	script	Fr	GDT, MS, Sample, TERMIUM
software	bibliothèque de logiciels	Fr	GDT, MS, Sample,

library			TERMIUM
software library	bibliothèque logicielle	Fr	TERMIUM
software library	librairie	Fr	Sample
software library	logithèque	Fr	GDT, Sample, TERMIUM
software library	software library	En	GDT, TERMIUM
spam	arrosage	Fr	GDT
spam	bulk e-mail	En	GDT
spam	courriel non sollicité, courriel de masse non sollicité	Fr	GDT, TERMIUM
spam	courriel indésirable	Fr	GDT
spam	courriel-poubelle	Fr	GDT
spam	courrier indésirable	Fr	MS
spam	junk mail, junk e-mail, electronic junk mail, junk electronic mail	En	GDT
spam	pourriel	Fr	GDT, TERMIUM
spam	spam	En	GDT, MS, Sample, TERMIUM
spam	UBE	En	TERMIUM
spam	UCE	En	GDT
spam	unsolicited e-mail, unsolicited bulk e-mail, unsolicited commercial e-mail	En	GDT, TERMIUM
super computer	méga-ordinateur	Fr	TERMIUM
super computer	super computer	En	GDT, TERMIUM
super computer	super ordinateur	Fr	GDT, MS, Sample, TERMIUM
super computer	super-calculateur	Fr	GDT, TERMIUM
super computer	super-scale computer	En	TERMIUM
system monitor	moniteur	Fr	TERMIUM
system monitor	moniteur système, moniteur de système	Fr	MS, Sample
system monitor	system monitor	En	GDT, MS, TERMIUM
system monitor	système d'auto-contrôle	Fr	TERMIUM
system monitor	système d'auto-surveillance	Fr	TERMIUM
system monitor	système de surveillance	Fr	GDT, TERMIUM
text processing	text processing	En	GDT, TERMIUM

text processing	traitement texte, traitement de texte	Fr	GDT, MS, Sample, TERMIUM
text processing	traitement des mots	Fr	TERMIUM
text processing	word processing	En	GDT, TERMIUM
text processing	WP	En	TERMIUM
touchpad	bloc à effleurement	Fr	GDT, TERMIUM
touchpad	dispositif de pointage tactile	Fr	TERMIUM
touchpad	pavé tactile	Fr	GDT, Sample, TERMIUM
touchpad	pointeur tactile	Fr	GDT
touchpad	touchpad, touch-pad	En	GDT, Sample, TERMIUM
touchpad	touch-sensitive pointing device	En	TERMIUM
touchpad	trackpad, track-pad	En	GDT, TERMIUM
update manager	gestionnaire de mise à jour, gestionnaire des mises à jour	Fr	MS, Sample
update manager	update manager	En	MS
user	usager	Fr	Sample, TERMIUM
user	user	En	GDT, TERMIUM
user	utilisateur, utilisatrice	Fr	GDT, Sample, TERMIUM
web page	page internet	Fr	Sample
web page	page web	Fr	GDT, MS, TERMIUM
web page	web page	En	GDT, TERMIUM

APPENDIX 3: Raw data - Terminometric analysis

Overarching concept	Term	IC	Source(s)	# of occ.	Pattern(s)
blog	blog	0.54	GDT, MS, Sample, TERMIUM	50	Dir. borr.
blog	blogue	0.45	GDT, Sample, TERMIUM	41	Adp. borr.
blog	carnet web	0.01	GDT, TERMIUM	1	Adp. borr., Compound, Use of ex. res.
blog	cybercarnet	0.00	GDT, TERMIUM	0	Deriv./affix., New creation
button	bouton	0.99	GDT, MS, Sample, TERMIUM	9999 *	Use of ex. res.
button	touche de fonction	0.01	TERMIUM	1	Compound, Use of ex. res.
button	bouton poussoir, bouton-poussoir	0.00	TERMIUM	0	Compound, Deriv./affix., Use of ex. res.
button	bouton de fonction	0.00	TERMIUM	0	Compound, Use of ex. res.
command	commande	0.98	GDT, MS, Sample, TERMIUM	770	Use of ex. res.
command	command	0.01	GDT, TERMIUM	8	Dir. borr.
command	instruction	0.01	TERMIUM	9	Deriv./affix.
cursor	curseur	1.00	GDT, MS, Sample, TERMIUM	25	Use of ex. res.
cursor	marqueur	0.00	TERMIUM	0	Deriv./affix.
default	par défaut	0.98	GDT, MS, Sample, TERMIUM	9999 *	Compound, Use of ex. res.
default	default	0.01	TERMIUM	1	Dir. borr.
default	implicite	0.01	TERMIUM	1	Use of ex. res.
disk space	espace disque	1.00	GDT, MS, Sample, TERMIUM	32	Compound, Use of ex. res.
display	affichage	0.97	GDT, Sample, TERMIUM	114	Deriv./affix.

display	visualisation	0.03	GDT, TERMIUM	4	Deriv./affix.
display	affichage numérique	0.00	GDT	0	Compound, Deriv./affix.
download	télécharger, téléchargement	0.87	GDT, MS, Sample, TERMIUM	69	Deriv./affix., New creation
download	download	0.13	GDT, MS, Sample, TERMIUM	10	Dir. borr.
e-mail	courriel	0.80	GDT, MS, Sample, TERMIUM	381	Compound, Compression, New creation
e-mail	e-mail, email	0.15	GDT, MS, Sample, TERMIUM	69	Dir. borr.
e-mail	mail	0.04	GDT, Sample	18	Dir. borr.
e-mail	courrier électronique	0.01	GDT, MS, TERMIUM	6	Compound, Deriv./affix.
e-mail	courriél	0.00	GDT	0	Compound, Compression, New creation
e-mail	CÉ	0.00	TERMIUM	0	Compound, Compression, New creation
e-mail	message électronique	0.00	GDT, TERMIUM	1	Compound, Use of ex. res.
e-mail	mél	0.00	GDT	0	Compound, Compression, New creation
internet provider	fournisseur internet	0.57	GDT, MS, TERMIUM	8	Compound, Deriv./affix.
internet provider	fournisseur d'accès internet, fournisseur d'accès à internet	0.14	GDT, MS, TERMIUM	2	Compound, Deriv./affix., Use of ex. res.
internet provider	ISP	0.07	GDT, TERMIUM	1	Dir. borr.
internet provider	fournisseur d'internet	0.07	Sample	1	Compound, Deriv./affix.
internet provider	FAI	0.07	GDT, TERMIUM	1	Compound, Compression, New creation

internet provider	prestataire de service	0.07	GDT	1	Compound, Deriv./affix., Use of ex. res.
internet provider	fournisseur d'accès	0.00	GDT	0	Compound, Deriv./affix.
internet provider	fournisseur de services internet	0.00	GDT, MS, TERMIUM	0	Compound, Deriv./affix., Use of ex. res.
internet provider	FSI	0.00	GDT, TERMIUM	0	Compound, Compression, New creation
internet provider	prestataire d'accès internet, prestataire d'accès à internet	0.00	TERMIUM	0	Compound, Deriv./affix., Use of ex. res.
internet provider	prestataire de services internet	0.00	GDT	0	Compound, Deriv./affix., Use of ex. res.
internet provider	prestataire internet	0.00	TERMIUM	0	Compound, Deriv./affix.
interoperability	interopérabilité	1.00	GDT, MS, Sample, TERMIUM	11	Deriv./affix.
interoperability	interfonctionnement, aptitude à l'interfonctionnement	0.00	GDT, TERMIUM	0	Deriv./affix.
keyboard	clavier	1.00	GDT, MS, Sample, TERMIUM	9999 *	Use of ex. res.
light	léger, légère	0.88	MS, Sample	21	Use of ex. res.
light	light	0.13	MS, Sample	3	Dir. borr.
log	log	0.65	Sample	39	Dir. borr.
log	syslog	0.13	TERMIUM	8	Dir. borr.
log	journal, journal système, journal de marche, journal d'exploitation	0.13	Sample, TERMIUM	8	Compound, Use of ex. res.
log	registre	0.08	MS	5	Use of ex. res.
login	connexion, se connecter	0.38	TERMIUM	49	Use of ex. res.
login	ouverture de session, ouvrir une session	0.32	GDT, Sample, TERMIUM	41	Compound, Deriv./affix., Use of ex. res.

login	login, log in, log-in	0.31	GDT, Sample, TERMIUM	40	Dir. borr.
login	entrée dans le système	0.00	TERMIUM	0	Compound, Use of ex. res.
memory	mémoire	0.99	GDT, MS, Sample, TERMIUM	170	Use of ex. res.
memory	storage	0.01	GDT, TERMIUM	1	Dir. borr.
memory	unité de stockage	0.01	TERMIUM	1	Compound, Deriv./affix., Use of ex. res.
memory	dispositif de stockage	0.00	TERMIUM	0	Compound, Deriv./affix.
microprocessor	processeur	0.62	GDT, Sample	93	Deriv./affix.
microprocessor	CPU	0.36	Sample	54	Dir. borr.
microprocessor	microprocesseur	0.02	GDT, MS, Sample, TERMIUM	3	Deriv./affix., New creation
motherboard	carte mère, carte-mère	0.69	GDT, MS, TERMIUM	22	Compound, Loan tran., Use of ex. res.
motherboard	carte maîtresse	0.19	Sample, TERMIUM	6	Compound, Use of ex. res.
motherboard	motherboard	0.13	GDT, TERMIUM	4	Dir. borr.
motherboard	carte principale	0.00	TERMIUM	0	Compound, Use of ex. res.
motherboard	carte de connexion	0.00	TERMIUM	0	Compound, Deriv./affix., Use of ex. res.
multiboot	multiboot, multi-boot	1.00	Sample	5	Dir. borr.
multiboot	amorçage multiple	0.00	GDT	0	Compound, Deriv./affix.
multiboot	démarrage multiple	0.00	MS	0	Compound, Deriv./affix.
office suite	suite bureautique, suite de bureautique	0.51	GDT, TERMIUM	35	Compound, Deriv./affix., Use of ex. res.
office suite	environnement de bureau	0.30	Sample	21	Compound, Use of ex. res.
office suite	suite Office	0.10	MS	7	Adp. borr., Compound, Use of ex. res.

office suite	suite office	0.06	Sample	4	Adp. borr., Compound
office suite	environnement bureautique	0.03	Sample	2	Compound, Deriv./affix.
online	en ligne	0.97	MS, Sample, TERMIUM	127	Compound, Use of ex. res.
online	connecté, en connecté	0.02	GDT, TERMIUM	3	Compound, Use of ex. res.
online	online, on-line	0.01	TERMIUM	1	Dir. borr.
online	en directe	0.00	TERMIUM	0	Compound, Use of ex. res.
operating system	système d'exploitation	0.53	GDT, MS, Sample, TERMIUM	153	Compound, Deriv./affix., Use of ex. res.
operating system	OS	0.45	MS, Sample, TERMIUM	130	Dir. borr.
operating system	SE	0.02	GDT	6	Compound, Compression, New creation
operating system	logiciel d'exploitation	0.00	GDT	0	Compound, Deriv./affix.
panel	panneau	0.91	MS, Sample	41	Use of ex. res.
panel	panel	0.09	TERMIUM	4	Dir. borr.
panel	écran utilitaire	0.00	TERMIUM	0	Compound, Use of ex. res.
partition (drive)	partition	0.99	GDT, MS, Sample, TERMIUM	646	Deriv./affix.
partition (drive)	partitionnement	0.01	Sample, TERMIUM	8	Deriv./affix.
partition (drive)	division	0.00	TERMIUM	0	Deriv./affix.
programming	programmation	0.84	GDT, MS, Sample, TERMIUM	90	Deriv./affix.
programming	codage	0.16	GDT	17	Deriv./affix.
remote desktop	bureau à distance	0.83	MS, Sample	25	Compound, Use of ex. res.
remote desktop	bureau distant	0.10	Sample	3	Compound, Loan tran., Use of ex. res.
remote desktop	remote desktop	0.07	MS	2	Dir. borr.
removable	amovible	1.00	MS, Sample	6	Deriv./affix.
removable	détachable	0.00	GDT, TERMIUM	0	Deriv./affix.

removable	démontable	0.00	GDT, TERMIUM	0	Deriv./affix.
removable	supprimable	0.00	TERMIUM	0	Deriv./affix.
script	script	1.00	GDT, MS, Sample, TERMIUM	9999 *	Use of ex. res.
script	scénario	0.00	TERMIUM	0	Use of ex. res.
software library	logithèque	0.54	GDT, Sample, TERMIUM	83	Compound, Compression, New creation
software library	librairie	0.45	Sample	69	Adp. borr.
software library	bibliothèque de logiciels	0.01	GDT, MS, Sample, TERMIUM	1	Compound, Deriv./affix., Loan tran.
software library	bibliothèque logicelle	0.00	TERMIUM	0	Compound, Deriv./affix.
spam	spam	0.83	GDT, MS, Sample, TERMIUM	24	Dir. borr.
spam	pourriel	0.17	GDT, TERMIUM	5	Compound, Compression, New creation
spam	courriel non sollicité, courriel de masse non sollicité	0.00	GDT, TERMIUM	0	Compound, Compression, New creation, Use of ex. res.
spam	courriel indésirable	0.00	GDT	0	Compound, Deriv./affix., New creation
spam	courriel-poubelle	0.00	GDT	0	Compound, New creation, Use of ex. res.
spam	arrosage	0.00	GDT	0	Deriv./affix.
spam	courrier indésirable	0.00	MS	0	Compound, Deriv./affix., Use of ex. res.
super computer	super ordinateur	0.83	GDT, MS, Sample, TERMIUM	10	Compound, Deriv./affix.
super computer	super-calculateur	0.17	GDT, TERMIUM	2	Deriv./affix., New creation
super computer	méga-ordinateur	0.00	TERMIUM	0	Deriv./affix., New creation

system monitor	moniteur système, moniteur de système	1.00	MS, Sample	14	Compound, Use of ex. res.
system monitor	système de surveillance	0.00	GDT, TERMIUM	0	Compound, Deriv./affix., Use of ex. res.
system monitor	système d'auto-surveillance	0.00	TERMIUM	0	Compound, Deriv./affix., Use of ex. res.
system monitor	système d'auto-contrôle	0.00	TERMIUM	0	Compound, Deriv./affix., Use of ex. res.
system monitor	moniteur	0.00	TERMIUM	0	Use of ex. res.
text processing	traitement texte, traitement de texte	1.00	GDT, MS, Sample, TERMIUM	16	Compound, Deriv./affix., Use of ex. res.
text processing	traitement des mots	0.00	TERMIUM	0	Compound, Deriv./affix., Use of ex. res.
touchpad	pavé tactile	0.53	GDT, Sample, TERMIUM	8	Compound, Deriv./affix., Use of ex. res.
touchpad	touchpad, touch-pad	0.27	GDT, Sample, TERMIUM	4	Dir. borr.
touchpad	trackpad, track-pad	0.20	GDT, TERMIUM	3	Dir. borr.
touchpad	bloc à effleurement	0.00	GDT, TERMIUM	0	Compound, Deriv./affix., Use of ex. res.
touchpad	dispositif de pointage tactile	0.00	TERMIUM	0	Compound, Deriv./affix.
touchpad	pointeur tactile	0.00	GDT	0	Compound, Deriv./affix.
update manager	gestionnaire de mise à jour, gestionnaire des mises à jour	0.78	MS, Sample	29	Compound, Deriv./affix.
update manager	update manager	0.22	MS	8	Dir. borr.
user	utilisateur, utilisatrice	0.89	GDT, Sample, TERMIUM	820	Deriv./affix.
user	usager	0.10	Sample, TERMIUM	93	Deriv./affix.
user	user	0.01	GDT, TERMIUM	11	Dir. borr.

web page	page web	0.95	GDT, MS, TERMIUM	72	Adp. borr., Compound, Use of ex. res.
web page	page internet	0.05	Sample	4	Compound, Use of ex. res.

*: Not the actual number of occurrences. See Footnote 49, p. 68.