

Toward a corpus-based multilingual terminology database for Intercultural Communication

María Iglesias Vázquez, Charlotte Venema, Marie Steffens

Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands

E-mail: mariaiglesiasv00@gmail.com, cacvenema@gmail.com, m.g.steffens@uu.nl

Abstract

This contribution focuses on the methodological aspects of the ICoMuTe project aiming to design a corpus-based multilingual terminology database for Intercultural Communication (ICC). The project seeks to explore how ICC terms relate to each other within six European languages (Dutch, English, German, French, Italian, Spanish), how these terms are connected to their scientific and cultural contexts, and how they can be translated across different languages and cultures while preserving meaning.

The selected approach is corpus-based, using comparable corpora of ICC handbooks and a parallel corpus of texts produced by the European Parliament dealing with key questions related to ICC. Using text recognition and data mining tools (e.g., Sketch Engine), the most frequent ICC terms per language are extracted and analysed in context. To account for the culturally specific aspects of terms while achieving a high degree of cultural neutrality, a semantic model based on tags has been developed for comparing and linking terms across languages in a neutral manner, but natural language corpus-based definitions are also provided that reflect the cultural load of each term.

The main findings suggest that semantic tags are relevant to balance the cultural specificity and neutrality of ICC terms, and that English acts as a reference linguistic and cultural framework for the emergence and development of terms in other languages.

Keywords: intercultural communication; multilingual terminology; corpus-based lexicography; lexical functions; semantic primes.

1. Introduction

Intercultural communication (ICC) has become a crucial area of study in an increasingly globalised world, where interactions across linguistic and cultural boundaries are more frequent than ever. This means that understanding how cultural contexts shape language and meaning is essential for effective communication. The complexity of ICC is reflected in the ways its specialised terminology can convey culturally specific meanings that differ across languages, so that even seemingly equivalent terms may carry distinct connotations or underlying conceptual frameworks in different linguistic communities. This highlights the necessity of exploring the semantic properties of ICC terminology across languages to enhance clearer and more effective cross-cultural understanding.

One of the primary challenges in multilingual ICC research is the lack of a standardised, comprehensive glossary. Even though various multilingual databases exist in other specialised fields, the absence of a structured, cross-linguistic approach to ICC terminology creates difficulties for professionals and learners who rely on precise definitions to navigate intercultural interactions and research. As a result, the ICoMuTe project aims to construct a corpus-based multilingual glossary and to investigate to which extent ICC terms convey scientific cultural meaning specific to six European and world-languages (Dutch, English, German, French, Italian, and Spanish).

The present research is both scientifically and socially relevant, meeting the need to collect and represent the metaknowledge of the field, and to create a baseline for research and teaching for optimal cross-culture understanding of terms in different languages.

This contribution will focus mainly on methodological aspects related to the design of the database but will also present some insightful concrete examples of cross-cultural differences between terms in the different languages.

2. Terminological definitions from a sociocognitive perspective

Terms designate specialised knowledge and differ from general-language words (Adelstein, 2004), especially in pragmatic and communicative aspects. Consequently, although they are part of a language's general system, they also form specific subsystems. One of their key characteristics, according to Cabré (1999), is that terms “are much more closely related to the other terms in the same discipline, with which they form specific structures” (p. 82), underlining semantic interconnection and interdependence.

Theoretical constructs from Cabré's communicative theory of terminology (1999), Temmerman's sociocognitive theory (2000), and Diki-Kidiri's cultural terminology (2022) have been used and adapted for the development of the ICoMuTe project. Cabré (1999) diverges from earlier prescriptive traditions and asserts that terminology's main concern are specialised languages, defined as linguistic subcodes varying by “subject field, type of interlocutors, situation, speaker's intentions, context, etc.” (1999: 59). By focusing on pragmatic and communicative dimensions, Cabré makes space for studying variation in terminology. She notes that specialised language varies in the same fashion as general language because terms can be used to transmit specialised knowledge between individuals with different expertise levels, triggering contextual variation.

Similarly, Temmerman's (2000) sociocognitive approach issues from the idea that expert groups are heterogeneous, and specialised fields are dynamic and interdisciplinary. Thus, synonymy and polysemy are considered in terminology due to their functionality, accommodating language evolution in expert domains. Real-world changes may alter mental prototypes and categorisation, sometimes clashing with

language constraints and prompting variation in lexical units (Fuertes-Olivera & Tarp, 2014).

In line with Cabré and Temmerman, developers of cultural terminology argue that terms and concepts are always culture-bound (Diki-Kidiri, 2022). Culture conditions how people and communities perceive and categorise reality and therefore also impacts the formation of specialised knowledge and terms. Cultural terminology considers that terms are “the cultural expression of specialized knowledge in any natural language belonging to a community of speakers” (Diki-Kidiri, 2022: 200). This approach foregrounds the human dimension of terminology, given that humans are the agents constantly generating and integrating new knowledge and experience. Speaker communities thus become the authority agents in cultural terminology, since they are shaped by their unique history, habits, relationships, and communicative practices. Culture, in fact, “shapes the totality of a person’s relationship with whatever exists” (Diki-Kidiri, 2022, p. 203), including linguistic signs.

Multilingual terminology must consider these principles to manage diverse linguistic and cultural contexts, and to develop terminological resources. Three reasons support this proposal: first, different languacultures (Agar, 1994) may lack equivalent experiences or knowledge, leading to the appropriation or adaptation of others’; second, specialised communities are heterogeneous across languacultures, so some terms may vary or not exist depending on context; and third, specialised fields are communities of speakers themselves, open to new experiences and knowledge.

Terms can be studied from formal (designation), semantic (concept), and functional (pragmatic) perspectives (Cabré, 1999). Even though the formal and functional dimensions are considered in the ICoMuTe project, the main focus is on the semantic dimension of terms. Semantically, concepts are mental constructs representing elements of internal or external reality and involving some level of abstraction (ISO 704, 2022). However, the categorisation of these concepts by individuals and communities is culturally influenced (Diki-Kidiri, 2022). The cultural terminology lens helps explain the presence or absence of certain terms in specified fields across languages.

According to ISO standard 1087 (2019), terminological definitions are “statements which describe a concept, and which permit its differentiation from other concepts within a conceptual system”. Terminological definitions illustrate how concepts are understood and expressed in special communication. They are typically composed of three elements: domain, initial definator, and characteristics (Vézina, 2009). Since all terms considered in the project belong to ICC, the domain element is not discussed here. Initial definitors are followed by term characteristics, that can be essential, i.e. common to the whole conceptual group; or complementary, i.e. non-relevant, additional traits (Cabré, 1999). Following Wierzbicka (1985), definitions are also exhaustive, offering a full explanation of the concept, and reductive. Reductive definitions explain concepts “via simpler, and not via more or equally complex concepts” (Wierzbicka,

1985: 45). The innovative semantic model developed to allow the comparison of ICC terms across languages is presented in Section 4.

From a sociocognitive perspective, the creation of our multilingual ICC terminological database meets the following objectives:

1. Exploring how ICC terms can be related to each other in semantic networks within individual language systems.
2. Investigating how the concepts associated to ICC terms are related to the scientific culture they belong to.
3. Researching how terms and their semantic relationships can be accurately translated into other languages while considering their diverging scientific cultures.

3. A corpus-based approach to identify ICC terms

For the exploration and extraction of ICC terminology, a corpus-based approach has been adopted to create a multilingual database in six European or world languages: Dutch, English, German, Spanish, French, and Italian. Language selection was based on the language offer of the Master's degree in Intercultural Communication at Utrecht University, which includes these six languages.

The created corpora are classified into a comparable handbook corpus and a parallel institutional documentation corpus. The handbook corpus includes manuals of intercultural communication natively written in the six languages, at least two per language. Manuals were chosen because they aim to offer clear and comprehensive insights into subject fields, usually assuming the reader's lack of expertise, which means that they can provide rich conceptual contexts to define concepts. Even if the most recent works or editions have been preferred, the constitution of the corpus is dependent on the editorial liveliness in the different languages: there are obviously more recent manuals published in English than in Italian. The constitution of the corpus depended on opportunity (existence and availability of manuals). In a second phase of the project, it could be extended.

The institutional documentation corpus includes 12 parallel texts selected from the European Parliament repository. The search in the EP repository was filtered by period (2014-2019 and 2019-2024) and keyword (*intercultural*). The EP repository contains parallel translations in all the languages of the EU, which provided us with access to the same text in the six languages of the project. The main purpose of this corpus was to observe how the selected terms of the glossary were used and to extract textual context. Since the results were filtered through the keyword *intercultural*, all texts deal, to a greater or lesser extent, with issues related to interculturality and intercultural communication, and how these are regularised, promoted, and protected within the EU

framework. This corpus-based approach accounts for the textual (occurrence in real documents), conceptual (characteristics of concepts and their relations to others) and communicative (context) dimensions of terms.

Terminological extraction was accomplished with the aid of Sketch Engine. The handbook corpora in each language were compiled and managed through the software to produce comprehensive lists of single- and multi-term keywords. For that purpose, the Sketch Engine in-built reference corpora of each language were used for contrast. Stopword lists were applied to filter out highly frequent but non-informative words, and lemmatization was used to group different forms of the same word, ensuring a more accurate frequency analysis.

Sketch Engine produced two first lists of 150 single- and multi-term keywords in the six languages. Each of these lists was then individually examined to produce final lists of about 50 single- and multi-terms. Through observation of frequency and occurrence of terms, criteria of exclusion were established to select or rule out terms (e.g., terms were too broad or narrow, or they referred to theories, disciplines, or approaches). However, not all terms were selected purely based on frequency. Some terms were included because of their suitability, meaning that even if they did not rank highly in frequency, they occurred often enough to allow extraction of a meaningful definition based on context. Conversely, some highly frequent words were excluded when their occurrences were largely non-informative, such as when they appeared mainly in titles, headings, or other repetitive structural elements. The motivation behind this decision was that the final glossary ought to be a foregrounding, introductory tool, with the potential to be expanded, but which should initially account only for fundamental ICC concepts.

The 50-term mark was easily reached in the English corpus, but corpora in the five other languages yielded less terms that adhered to our exclusion/inclusion criteria. Consequently, those lists were carefully scrutinised through the search engine of the terminological software to access terms that, despite being less frequent, had potential to be included in the final lists.

The selected terms in the six languages can be classified into three categories:

1. Potentially equivalent terms that were frequent in all (or most) languages: some core terms of ICC appeared frequently in all or most languages, such as *ethnocentrism*, *intercultural contact*, *intercultural competence*, *interculturality* or *stereotype*, among others. These terms have the potential to explore language- and culture-specific definitions, show the interconnection between specific scientific cultures, and assess the applicability of semantic features across languages.
2. Frequent terms in one language that were significantly less frequent in others: some relatively frequent (and therefore significant) terms in one language occurred less frequently in others. Examples were the English *cultural group*, the

Spanish *tercera cultura*, the German *Adaptation*, the Dutch *culturele synergie*, the French *identité nationale*, and the Italian *sensibilità interculturale*.

3. Relatively frequent terms in one language what were absent in (almost) all others: some frequent terms in one language were missing or could not be found in the corpora of other languages. Examples include the English *small talk*, the Spanish *gestualidad* (also found in the Italian corpus *gestualità*), the German *Amerikanisierung*, the Dutch *gezichtsverlies*, the French *distance sociale*, and the Italian *espressività emotive*.

Terms belonging to groups 2 and 3 were selected because they could expose the cultural ethnocentrism of ICC. While Goddard (2004) suggests that ethnocentrism should be erased from the conceptual content of terms, this is challenged when terms themselves are culturally dependent as products of the scientific cultural context in which they occur. The absence in one language of terms that frequently appear in another might be a sign of cultural load in terminology and scientific disciplines. Despite being developed within the semi-controlled context of specialised language, terms still result from scientific communities of practice that are influenced by their sociocultural circumstances. For instance, the occurrence of *gestualidad* in the Spanish corpus and *gestualità* in the Italian corpus but the absence of equivalents in other corpora might expose something about the community in which Spanish and Italian ICC are embedded, and which produces specific terminological needs. While this does not signal that body language / gesture is an underdeveloped field in other languages, it might be an indicator of each specific languaculture's priorities. The fact that body language appears as a concept in the Spanish and Italian ICC handbooks, and therefore its terminology, reflects ethnocentric tendencies in the study of specialised language.

4. Corpus-based definitions

If possible, terminological definitions were directly extracted from the definitional context provided by the ICC handbooks. When this was not possible, they were either (1) produced *ad hoc* based on the specialised knowledge identified in the corpora, or (2) searched for in specialised documentation other than the corpora.

Concept definition writing was based on Remígio's (2013) guidelines, which consist of "generation of concordances; identification of contexts rich in conceptual information; identification and highlighting of conceptual characteristics from those contexts; systematisation and selection of the conceptual characteristics identified; and definition writing itself" (p. 204). Concordances were generated using the 'Concordance' tool in Sketch Engine, which allows to access all the textual contexts in which a single- or multi-term appears. This facilitated the identification of contexts rich in conceptual information, which were crucial to find and select conceptual characteristics in a reliable and quick manner.

The complexity of this process can be illustrated by the Spanish single-term *estereotipo*. Several encyclopaedic and semi-definitional contexts were identified in Alsina (1999) out of which the following were extracted:

Los estereotipos nos permiten encasillar, tipificar de manera fácil a todo el mundo. [...] Recurrimos a los estereotipos como generalizaciones simplistas que forman parte de un cierto sentido común o comunitario de una cultura determinada. [...] Muchos estereotipos están cargados de connotaciones negativas. Así mediante el estereotipo se alimenta la negatividad de “el otro”. (p. 75)¹

Following Cabré’s (1999) and Vézina’s (2009) guidelines for terminological definition writing, our proposed definition for *estereotipo* is:

estereotipo: Generalización simplista que permite encasillar y tipificar de manera fácil a los individuos de un grupo social, generalmente de modo negativo.²

The genus remotum *generalización* was chosen as initial definator for the term, and the essential characteristics of the concept identified in the semi-definitional and encyclopaedic contexts were chosen and adapted for the definition.

The second process of writing definitions involved the search and selection of other specialised documentation for terms for which the corpora did not provide enough definitional context. In the case of terms in English, this resulted a smooth endeavour due to the extensive existing literature in the field of intercultural communication. However, this process posed a greater challenge in all other languages. There seems to be a loophole at the level of handbooks in intercultural communication and cultural studies in languages other than English. In most cases, it was necessary to resort to highly specific articles that provided conceptual definitions, and those definitions were often translations from an English source.

This scientific gap seems to be suggestive of a sort of cultural imperialism within the field of intercultural communication. ICC research in English, if considered a scientific cultural process that results in cultural products such as theories, approaches and methodologies, which are then gathered in handbooks, impacts and determines the (under)development of the field within other languages and cultures. In fact, this broader issue could risk the development of a truly ethnorelative semantic model in the cases in which terminographical definitions in one language have been produced based on definitions rooted in the Anglophone tradition. The problem goes a step further in

¹ “Stereotypes allow us to categorise, to easily typify everyone. [...] We resort to stereotypes as simplistic generalisations that are part of a certain common or communal sense of a given culture. [...] Many stereotypes are loaded with negative connotations. Thus, through stereotyping, the negativity of ‘the other’ is nurtured.”

² *Stereotype*: Simplistic generalisation that allows individuals in a social group to be easily categorised and typified, usually in a negative way.

languages such as Italian and German. Their handbooks often use ICC terms in English, sometimes even without providing language-specific translations, and therefore anchor their associated concepts in the English-speaking tradition. This was the case with *critical incidents* in German, and *worldview* and *cultural framework* in Italian, for instance. This virtual shortage of specific and original ICC handbooks, and sometimes terminology, in languages other than English is noteworthy. It highlights the pressing need for a multilingual ICC glossary that aims to bridge terminological gaps and alleviate the presumed hierarchy among different language-specific ICC research.

This goal was pursued by writing conceptual definitions in the language of the term that was being defined. This does not imply that terms in each language remained isolated and untranslatable; on the contrary, it facilitated the search or production of functional equivalents, thanks to exhaustive definition.

5. A dedicated semantic model

In order to explicit cross-linguistic connections, a semantic model based on the terminological definitions of the concepts was constructed. This model constituted a tag system, and tags, i.e. free keywords related to lexical items (Vuorikari, 2009), were then applied to each selected term. For a full account of the developed tag system, see Appendix A.

The two main theoretical sources that constitute the basis of the model are the natural semantic metalanguage (Goddard & Wierzbicka, 1994; 2002) and the lexical functions (Žolkovskij & Mel'čuk, 1965).

The natural semantic metalanguage (NSM), developed for lexicography, aimed at discovering the smallest set of basic concepts through which all other concepts could be explained, for simplicity and clarity in the definition of words belonging to general language, which its authors based on reductive paraphrase. This implies that “the meaning of any semantically complex word can be explicated by means of an exact paraphrase composed of simpler, more intelligible words than the original” (Goddard & Wierzbicka, 2002: 5). Those simpler words constitute the core vocabulary of a language and, according to research, have linguistic exponents in all languages (Goddard & Wierzbicka, 2007). Therefore, this proposed metalanguage, even if it is formulated in English, could be understood as a conceptual lingua franca that allows for the explanation of meaning across languages and cultures.

Understood as such, NSM becomes culturally neutral and useful to remove terminological ethnocentrism from conceptual definitions. In terminological ethnocentrism, complex and culture-specific words of one language or culture are used as descriptive elements to analyse the meanings and assumptions of another language or culture, imposing an outsider perspective (Goddard, 2004). Despite Goddard’s suggestion, languages cannot be regarded as culturally and pragmatically neutral even

in their core components. For that reason, we consider that conceptual definitions of terms should remain culturally dependent due to their belonging to a specific linguistic and scientific tradition. However, the means to connect those concepts cross-linguistically, namely, our proposed semantic model, should aim at ethnorelativism. Wierzbicka's (1985) insights into the creation of a NSM served as a starting point for our ICC semantic metalanguage, since the author suggests that NSM can be modified for specific purposes. Therefore, following the guidelines of NSM, an ICC semantic metalanguage in the form of a tag system can be created for the construction of subject-specific definitions.

From a more language-specific perspective, Žolkovskij and Mel'čuk (1965) introduced the concept of lexical function (LF). Lexical functions help to establish correspondences between lexical units in different languages because they indicate each lexical unit's function in relation to another (Kahane & Polguère, 2001). Consequently, any LF (which we name 'f') denotes a set of lexical units that are linked by a corresponding lexical relation, and which is realised in context. In their grounding work, Žolkovskij and Mel'čuk (1965) produced a set of around sixty primitive relations. Those relations follow the formula 'f[base1]=base2', meaning that the relation 'f' holds from 'base1' (keyword) to 'base2' (value of 'f').

To propose a semantic metalanguage through which ICC terms can be tagged and retrieved, we have combined the suggestions made by the NSM project and the ideas exposed by the LF theory. Some tags were drawn from NSM and LF denotations, and others were produced for the special needs of ICC.

Two relevant aspects under consideration in tag production were hierarchical and combinational patterns. The way in which tags interact with each other to produce definitions is ruled by a hierarchical structure, in which a set of basic-level tags are applied to concepts to provide their core characterisation. These include tags such as *feeling*, *process*, *event*, *situation*, *system* or *pattern*, among others. Other tags follow, adding complexity to the basic-level tag and relating to it hierarchically. Consequently, the order in which tags are associated to terms is relevant for their definition, retrieval, and cross-linguist comparison. There are also limitations as to how tags can be combined. Even though some tags denote full meaning by themselves and can be used in isolation, others hold a subordinate position in relation to others. Subordinate tags are the tags *kind* and *part*, based on and adapted from NSM, and all the LF-based tags, since LFs denote conceptual relations. Subordinate tags can be further classified into modifying and relational. Modifying tags are applied to only one other tag and they produce changes in its role, while relational tags denote a relation between two tags. Illustrative examples of how tags were applied and operationalised can be found in Appendix B, which includes a filled-in terminological record.

Challenges arose during the processes of tag selection and definition. This was especially relevant when there was potential overlap between the NSM- and LF-based tags, or

when an ICC-specific tag was close in meaning to one NSM- or LF-based tag. The following examples illustrate the cases of choosing one tag denomination over another, and the differences between closely related tags:

- Choosing *group* over *people*: NSM suggests *people* as a semantic primitive denoting more than one personal entity (Goddard & Wierzbicka, 1994). This tag candidate was ruled out due to its vagueness in denoting the links between subjects, in this case, between those that are considered to belong to one unit. On the contrary, the ICC-specific tag *group* was suggested, as it stresses that the individuals that form the unit find themselves in relation to one another, and potentially in contrast to more groups of individuals.
- Difference between *agent* and *individual*: personal entity and its diverse nuances are denoted by two closely related tags, namely, *agent* and *individual*. The tag *individual* simply indicates a single subject placed in contrast to a bigger, social group. However, definitions of concepts such as *intercultural competence* or *lingua franca* seem to consider more than the individual dimension of a personal entity. Opting to use a lingua franca or developing intercultural competences require active will and engagement from the involved individuals, causing them to perform some degree of agency. While the tag *individual* denotes a more ‘passive’ personal entity that is rather the object of external actions or events, the tag *agent* denotes an ‘active’ subject that performs those actions and events.

The production of tags also involved their clear and univocal definition. Since the goal of tag production was to create a set of tags that could be applied to all selected ICC terms and used equally across languages, tag definition had to serve a dual purpose. On the one hand, definitions had to remain partially broad to be applicable to as many terms as possible without restricting or conditioning their meaning. On the other hand, they had to aim at some level of specificity because they would be used for terminographical definition, which is per se specific. Consequently, tag definitions were mainly extracted from lexicographic works, that is, directly from their meaning and usage in the contexts of NSM and LF, and from dictionaries, namely, Cambridge Dictionary and Merriam-Webster Dictionary. However, most definitions were also adapted to express the nuances required by ICC terminology.

Tag definitions were essential for tag application, since they were the starting point to combine and assign tags to terms and produce functional descriptions of their concepts. Then, tag application helped to understand their interactional and combinatorial patterns, specially modifying and relational tags. For instance, the modifying tag *antimagn* follows the formula *antimagn[base]*. In this case, base is a concept, and *antimagn* modifies it to indicate that it is not intense. As such, in the combination *antimagn[understanding]* applied to the definition of *prejudice*, the modifying tag *antimagn* indicates that the concept *understanding* (mental grasp, comprehension) applies to a lesser or weakened extent to the definition of *prejudice*, since the holder or

prejudice forms their belief without evident reason. Another example of subordinate tag is the relational tag *manif*, which follows the formula *manif[base1, base2]*. In this case, *base1* is a concept that manifests itself in or is applied to *base2*, another concept. As such, in the combination *manif[feature, group]* applied to the definition of *essentialism*, the tag *manif* denotes that *essentialism* derives, partly, from a *feature* being manifested in or applied to a *group*.

Tag application also showed that even though modifying and relational tags could only remain so, there were cases that called for the usage of independent tags as either modifying or relational tags. This was mainly the case with tags derived from determiners, such as *same*, *other* or *own*, and adjectives, such as *common* or *different*. It could be that, since their corresponding lexical items in natural language usually function as modifiers, these tags can be readily transformed depending on the function they perform when defining a concept.

The produced tags were useful to establish relations between terms across languages. This was mainly achieved through those tags that were higher in the hierarchy, that is, those that express overarching meanings that often form the base of terminographical definitions. Despite the slight differences in the definitions of equivalent terms across languages, it was observed that these basic tags remained largely the same, which seems to reflect some level of cultural neutrality at the base of culturally specific scientific communities of ICC. What is more, some chains of tags that had been applied to concepts in one language were easily applicable to and functional for their equivalent concepts in other languages. This is illustrated by the term *socialisation* (en) and *socialización* (es), both of which were defined by the tag chain *process; individual; incep[perception]; incep[knowledge]; manif[culture, behaviour]; manif[culture, practice]; incep[belonging]*. In our tag system, the tag *incep* relates to a concept to indicate that it starts. Another example is the term *patrimonio cultural* (es) and *Kulturerbe* (de), both of which were defined by the tag chain *common[set]; plus[time]; exchange; group*. These cases suggest that the produced tag system was effective for language comparison.

However, there were also potentially equivalent terms in two languages that were assigned different tags, based on their culturally-specific definitions. An illustrative case is the term *culture shock* (en) and *choque cultural* (es). The Spanish definition of *choque cultural* lacks the emotional load that is heavy in the English definition of *culture shock*. Consequently, there is a significant difference between the identification of both terms, since *culture shock* is defined as a feeling and an experience, while *choque cultural* is defined as an event. This seems to have a relevant effect on how this concept is manifested and perceived in the two different languacultures: the English concept is individual-centred, focusing on the emotional response of the person undergoing culture shock, while the Spanish concept is context-centred, focusing on the disruption of known cultural parameters.

6. Treatment of synonymy and antonymy

Terminological records have been designed to systematically gather and present all the necessary terminological information for each selected term, following and adapting Cabré's (1999) guidelines. Appendix C includes the produced blueprint terminological record that was used for every terminological entry. Even though terminological record design and filling was fairly straightforward, slight challenges emerged when dealing with the sections 'Synonym', 'Variant' and 'Antonym'.

6.1 Synonymy and variation

Monolingual records were first created for each individual term in each language, and equivalents were then identified and paired. We differentiated perfect equivalents, terms in different languages that are suitable for each other, from quasi-equivalents, "terms which [mutually] refer to partially equivalent concepts out of context" (Diki-Kidiri, 2022: 215). Determining whether two terms were equivalents or quasi-equivalents in the different languages was mainly possible thanks to the produced tags system. We consider that synonyms are different denominations often used to refer to one same concept, but that might hold slight cognitive differences; while variants are denominative variations of one same concept due to dialectal, functional, discursive, or interlinguistic differences (Freixa, 2006).

We discuss the terminological record for culture shock as an example. This record includes culture confusion as a synonym. The underlying cognitive difference between both denominations is that some authors consider culture shock too negative, ignoring the learning potential of the experience, while culture confusion seems to describe the process of disorientation within a new culture in a more neutral way. The chosen entry is still culture shock because it occurs more frequently in the corpus (in fact, culture confusion does not occur in our corpus), despite its allegedly negative connotation. This example seems to showcase the usefulness of both the 'Note' and 'Synonym' sections, as they allow to stick to the criterion of frequency for term selection while accounting for problematic aspects related to the entry.

6.2 Antonymy

Similar to the identification of synonyms, monolingual records were first created for each individual term in each language. Subsequently, oppositional pairs were identified and analysed. Following Croft and Cruse (2004: 164–165), we consider antonyms to be lexical items that are inherently binary and "have a built-in twoness" (e.g. *good-bad*). Since we did not find such strict binary pairs in our corpus, we built on this perspective by considering the relation between lexical items that are presented by ICC scholars as opposed within a shared conceptual frame.

To illustrate this, we exemplify the terminological record *monochrome cultuur* and *polychrone cultuur* from the Dutch terminological records. The tags indicate that both terms share the same conceptual domain while differing in the direction of emphasis on time-related values. *Polychrone cultuur* contrasts with *monochrome cultuur* in that the former represents a flexible, multi-tasking approach to time, whereas the latter emphasizes structured, linear time management. Based on the contrast in time orientation and their shared conceptual frame, these terms function as conceptual antonyms.

7. Further development

Terminological records are not yet complete for all terms in all languages. This is the aim of phase 2 of the project. In the near future, however, searches in an online open database will enable the comparison of terms in different languages that correspond to the same tags, or differ only by some tags, in order to identify differences in connotations. This will be possible because all terminological records include clearly tagged data that is machine-readable and transposable to other online settings, which makes it adaptable to current developments in electronic terminography. The final platform will be based on a multilingual database of cross-referenced entries linked by systematically designed records that gather all relevant information for characterisation and retrieval. This database is intended to be extended and can be used for didactic purposes, but also for further research and more in-depth analyses of cross-cultural differences.

8. Acknowledgements

The authors want to thank the NWO, the Dutch Research Council, for making this project possible. They express also their gratitude to Marc Versluis from the Digital Humanities department at Utrecht University for his commitment to the online publication of the database.

9. References

- Adelstein, A. (2004). *Unidad léxica y valor especializado: estado de la cuestión y observaciones sobre su representación*. [Doctoral research]. Universitat Pompeu Fabra.
- Agar, M. (1994). *Language shock. Understanding the culture of conversation*. William Morrow.
- Alsina, M. R. (1999). *La Comunicación Intercultural*. Anthropos.
- Cabré Castellví, M. T. (1999). *Terminology: Theory, methods and applications*. John Benjamins Publishing Company.
- Cambridge Dictionary. (n.d.). *Cambridge Dictionary*. Retrieved April 1, 2025, from <https://dictionary.cambridge.org/>

- Croft, W. & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge University Press.
- Diki-Kidiri, M. (2022). Cultural Terminology. An introduction to theory and method. In P. Faber & M. L’Homme (Eds.), *Theoretical perspectives on Terminology: Explaining terms, concepts and specialized knowledge*, John Benjamins Publishing Company, pp. 197-226.
- Goddard, C. (2004). “Cultural scripts”: A new medium for ethnopragmatic instruction. In M. Achard, & S. Niemeier (Eds.), *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching*, Mouton de Gruyter, pp. 145-165.
- Goddard, C., & Wierzbicka, A. (1994). *Semantic and lexical universals – Theory and empirical findings*. John Benjamins Publishing Company
- Goddard, C., & Wierzbicka, A. (2002). *Meaning and Universal Grammar – Theory and empirical findings*. John Benjamins Publishing Company.
- Goddard, C., & Wierzbicka, A. (2007). Semantic primes and cultural scripts in language learning and intercultural communication. In F. Sharifian, & G. B. Palmer (Eds.), *Applied cultural linguistics: Implications for second language learning and intercultural communication*, John Benjamins Publishing Company, pp. 105-124.
- Freixa, J. (2006). Causes of denominative variation in terminology. *Terminology*, 12(1), pp. 51–77.
- Fuertes-Olivera, P. A., & Tarp. S. (2014). *Theory and practice of specialised online dictionaries: Lexicography versus Terminography*. DeGruyter.
- International Organization for Standardization. (2019). *Terminology work and terminology science — Vocabulary* (ISO standard No. 1087:2019).
- International Organization for Standardization. (2022). *Terminology work — Principles and methods* (ISO standard No. 704:2022).
- Merriam-Webster. (n.d.). *Merriam-Webster Dictionary*. Retrieved April 1, 2025, from <https://www.merriam-webster.com/>
- Remígio, A. R. (2013). The terminographical process: Phases and dimensions. *Meta*, 58(1), pp. 191–211.
- Temmerman, R. (2000). *Towards new ways of terminology description*. John Benjamins Publishing Company.
- Vézina, R. (2009). *La rédaction de définitions terminologiques*. Office québécois de la langue française.
- Vuorikari, R. H. (2009). *Tags and self-organisation: A metadata ecology for learning resources in a multilingual context* [Doctoral Thesis]. Open Universiteit.
- Wierzbicka, A. (1985). *Lexicography and conceptual analysis*. Karoma Publishers.
- Žolkovskij, A. K., & Mel’čuk, I. A. (1965). O vozmožnom metode i instrumentax semantičeskogo sinteza (On a possible method and instruments for semantic synthesis). *Naučno-tehničeskaja Informacija*, 5, pp. 23–28.

Appendices

Appendix A

Tag List

Table 1

Tags Produced from NSM

Determiners	Evaluations	Mental predicates	Speech	Taxonomy
same (m)	good (m)	thought	say	kind (r)
other (m)	bad (m)	knowledge	word	part (r)
		desire	true	
		feeling		

Table 2

Tags Produced from LFs

antimagn (m)
caus (r)
fact (m)
fin (m)
gener (m)
incep (m)
magn (m)
manif (r)
minus (m)
mix (m)
modif (m)
non (m)
plus (m)
real (r)

Table 3*ICC-specific Tags*

Subjects	Mental predicates	Actions	Events and situations	Theoretical accounts
agent	belief	action	common (m)	construct
culture	belonging	behaviour	communication	feature
ethnicity	consideration	challenge	conflict	pattern
group	experience	control	content	set
individual	fear	exchange	different (m)	system
language	goal	influence	event	tradition
race	idea	practice	problem	trend
relation	perception	process	situation	
self	understanding	speak	space	
own (m)				

Appendix B

Filled-in Terminological Record

Figure 1

Terminological Record of 'culture shock'

en	Record identifier
culture shock	
Source of term	
Hua, Z. (2019). <i>Exploring Intercultural Communication. Language in Action</i> . Routledge	
Grammatical category: noun	
Associated tag(s)	
feeling; experience; individual; non[same]; caus[culture, challenge]	
Definition	
Set of emotional reactions experienced by any individual who travels outside their home environment to another linguistic and cultural environment, and faces challenges adjusting to new cultural stimuli which have little or no meaning to them.	
Note	
Some scholars maintain that the term culture shock is too negative or extreme, ignoring the learning potential of this experience.	
Source of definition	
Adapted from Jackson, J. (2020). <i>Introducing language and intercultural communication</i> . Routledge	
Context	
The stress which occurs during <u>culture shock</u> very often results in 'cultural fatigue'. These terms reflect how people react when they find themselves relocated to a new culture and in a stage of transition – a process not unique to study abroad, but shared by many travellers and migrants.	
Source of context	
Hua, Z. (2019). <i>Exploring Intercultural Communication. Language in Action</i> . Routledge	
Synonym(s): culture confusion	Variant(s)
Antonym(s)	
Date: 02/04/2025	Author: María Iglesias Vázquez

Appendix C

Blank Terminological Record

Figure 2

Blank Terminological Record Used as Blueprint

en	Record identifier
Term entry	
Source of term	
Grammatical category	
Associated tag(s)	
Definition	
Source of definition	
Context	
Source of context	
Synonym(s)	Variant(s)
Antonym(s)	
Date	Author

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

