Document généré le 23 avr. 2024 23:25

Meta
Journal des traducteurs
Translators' Journal

META

Comparing across languages in corpus and discourse analysis: some issues and approaches

Charlotte Taylor et Dario del Fante

Volume 65, numéro 1, avril 2020

New Contexts in Discourse Analysis for Translation and Interpretation

URI : https://id.erudit.org/iderudit/1073635ar
DOI : https://doi.org/10.7202/1073635ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Taylor, C. & del Fante, D. (2020). Comparing across languages in corpus and discourse analysis: some issues and approaches. *Meta*, *65*(1), 29–50. https://doi.org/10.7202/1073635ar

Résumé de l'article

L'analyse du discours assistée par corpus est toujours, par définition, interdisciplinaire. Cela est particulièrement vrai lorsque l'on travaille sur la communication entre différentes langues assez lointaines les unes des autres ; on pénètre alors dans le domaine de la traductologie. L'objectif principal de cet article est d'offrir une réflexion à propos de la comparaison dans le contexte de corpus multilingues. Il en découle une série de questions clés comme celles-ci : comment aborder la comparaison entre corpus de langues différentes ou au sein d'un corpus multilingue ? Comment repérer les unités linguistiques significatives pour la comparaison dans ces contextes ? Comment savoir que l'on est en train de générer une comparaison d'éléments comparables ? Au moyen d'une série d'études de cas, notre réflexion commence par le plan lexical pour ensuite passer au plan sémantique, aux cadres du discours et aux traits rhétoriques. Grâce à ces questions et solutions toujours partielles, nous espérons entamer un débat productif qui alimentera cet ensemble de ressources.

Tous droits réservés © Les Presses de l'Université de Montréal, 2020

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

érudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Comparing across languages in corpus and discourse analysis: some issues and approaches

CHARLOTTE TAYLOR
University of Sussex, Brighton, United Kingdom
charlotte.taylor@sussex.ac.uk

DARIO DEL FANTE*
University of Padua, Padua, Italy
University of Sussex, Brighton, United Kingdom
dario.delfante@phd.unipd.it

RÉSUMÉ

L'analyse du discours assistée par corpus est toujours, par définition, interdisciplinaire. Cela est particulièrement vrai lorsque l'on travaille sur la communication entre différentes langues assez lointaines les unes des autres ; on pénètre alors dans le domaine de la traductologie. L'objectif principal de cet article est d'offrir une réflexion à propos de la comparaison dans le contexte de corpus multilingues. Il en découle une série de questions clés comme celles-ci : comment aborder la comparaison entre corpus de langues différentes ou au sein d'un corpus multilingue ? Comment repérer les unités linguistiques significatives pour la comparaison dans ces contextes ? Comment savoir que l'on est en train de générer une comparaison d'éléments comparables ? Au moyen d'une série d'études de cas, notre réflexion commence par le plan lexical pour ensuite passer au plan sémantique, aux cadres du discours et aux traits rhétoriques. Grâce à ces questions et solutions toujours partielles, nous espérons entamer un débat productif qui alimentera cet ensemble de ressources.

ABSTRACT

Corpus-assisted discourse studies is, by its nature, interdisciplinary. However, this need to reach across borders becomes even more salient when we study discourses across languages, and this represents a natural intersection with translation studies. The aim of this paper is to reflect on the issue of comparison in cross-linguistic corpus-assisted discourse studies, positing a series of key questions including: How do we compare across or within corpora containing different languages? How do we identify meaningful language units for comparison in this context? How do we know that we are comparing like with like? Using a series of case studies, we start by addressing how we can approach comparison at the lexical level. We then move on to consider methods which allow us to abstract above the lexical level using three case studies which illustrate the use of semantic fields, discourse frames and rhetorical features. By presenting some issues and partial solutions regarding comparison across and within multilingual corpora, we hope to initiate a productive discussion in which we will also be able to collectively enrich and inform this set of resources.

RESUMEN

El análisis de discurso asistido por corpus es siempre, y por naturaleza propia, interdisciplinario. Sin embargo, esta característica se hace especialmente evidente cuando nos acercamos a la comunicación a través de lenguas diversas, no siempre coincidentes con la original; al hacerlo, entramos de lleno en el ámbito de los estudios de traducción. El objetivo principal de este trabajo es reflexionar sobre la comparación propiciada por corpus multilingües, planteando una serie de preguntas clave, entre las que destacamos

las siguientes: ¿cómo propiciamos comparaciones entre corpus de lenguas diversas o en el seno de estos?; ¿cómo identificamos unidades lingüísticas significativas para la comparación en estos contextos?; ¿cómo sabemos que estamos generando comparaciones de componentes equiparables? Utilizando una serie de estudios de casos, nuestra reflexión comienza en el plano léxico para, luego, adentrarnos en el estudio de los campos semánticos, los marcos del discurso y los rasgos retóricos. Con este artículo, de preguntas y soluciones siempre parciales, esperamos iniciar un debate productivo con el que enriquecer colectivamente este conjunto de recursos.

MOTS-CLÉS/KEYWORDS/PALABRAS CLAVE

linguistique de corpus, analyse du discours, analyse critique du discours, analyse comparative du discours, comparaison interlinguistique
corpus linguistics, discourse analysis, critical discourse analysis, comparative discourse analysis, cross-linguistic comparison
lingüística de corpus, análisis del discurso, análisis crítico del discurso, análisis comparativo del discurso, comparación interlingüística

1. Introduction

This paper takes a reflexive approach to corpus-assisted analyses of discourse across languages. It follows in the tradition of Taylor and Marchi (2018) in trying to open up research processes by celebrating the impressive arsenal of tools we have available for our research, alongside a critical reflection on how we, as researchers, are influencing these research processes. This is a collaborative approach to discussing methodological innovation, which is important across all disciplines. However, in our experience, it is particularly important when we are working with and within new contexts, the focus of this special issue, and, specifically, when bringing in quantitative methodologies which may be seen as providing answers to questions of reliability and validity. In reality, of course, the findings are only as good as the design and so, while we have often found that corpus linguistics can offer answers on how to best to approach a research question, we have also found that it has raised many more. Taylor and Marchi (2018: 9) identify multilingual studies as an overlooked area within corpus and discourse studies and, they suggest, comparative studies also form a kind of triangulation if we are interested in identifying discourse patterns that transcend cultural and linguistic boundaries. Here, we examine some theoretico-methodological issues in comparing discourses across languages and we provide three case studies examining ways in which we can design meaningful units of analysis.

2. Cross-linguistic corpus-assisted discourse studies

2.1. Corpus-assisted discourse studies and "new contexts"

The combination of discourse analysis and corpus linguistics (as discussed in Baker 2006; Partington, Duguid, *et al.* 2013; Mautner 2015) draws upon different methodological traditions and theoretical assumptions about analysing language. As such, it is inherently open-minded as a methodology, well used to negotiating the tensions of qualitative and quantitative demands in research practice. From the corpus linguistic side, what we can gain is the ability to examine large numbers of texts

and, in so doing, to develop a different perspective on the data. As Fairclough (1989: 54) observed regarding the exertion of power by the media, "[a] single text on its own is quite insignificant: the effects of media power are cumulative, working through the repetition of particular ways of handling causality and agency, particular ways of positioning the reader, and so forth." Corpus linguistics provides a way of tracking this cumulative nature and this is particularly relevant in analysing collocations; that is the association between words. Knowing which words tend to go together can tell us more about the contextual meanings (including evaluative potential) of the lexical item we are particularly interested in. The (critical) discourse studies perspective, the other half of a corpus and discourse study, offers both a theoretical and methodological contribution to the framework. On the theoretical side, we have a set of assumptions about how language works in our society and the dialectical nature of that relationship which makes discourse analysis so relevant. On the methodological side, we have a collection of tools for analysis which allow us to identify and name patterns of language use.

The resulting combination, often referred to as corpus-assisted discourse studies or CADS, for short, has frequently been applied in interdisciplinary and multidisciplinary contexts. As Ancarno (2018) details, the interactions may come about 1) where the CADS work informs or is informed by work from other disciplines, 2) where the CADS work informs work outside academia, and 3) where CADS work is used in synergy with other disciplines. Within the context of this special issue, it is naturally the third of these that most concerns us. Our aim is to better describe the sub-area of cross-linguistic CADS and to reflect on some of the difficulties inherent in researching discourse across languages so that a dialogue may be opened with Translation Studies.

Although the term *cross-linguistic corpus-assisted discourse studies* (CL-CADS) is relatively recent (first used in Partington, Duguid, *et al*. 2013), the practice of working across languages using corpora has been going on for much longer. Furthermore, CL-CADS draws on the rich traditions of comparative discourse studies, comparative rhetorical studies and is necessarily informed by translation studies. However, until Vessey (2013) and Taylor (2014), there had been little reflection on the processes and methodological implications of examining discourses across languages when analysing corpora. Given the heterogeneity of approaches within just discourse analysis, it is to be expected that CL-CADS is not a systematic discipline endowed with a defined methodology and yet there are commonalities we can draw out.

2.2. Types of cross-linguistic corpus-assisted discourse studies

As discussed in Taylor (2014), the works which employ a cross-linguistic corpus-assisted discourse studies approach can be grouped in four categories, which differ according to the level of linguistic analysis, from a focus on the language itself to a more general cultural comparative interest, to cases where there is no comparative intent, but the datasets necessitate a multilingual approach.[1]

2.2.1. Language difference/similarity constitutes the object of study

The first category is constituted by those studies which are mainly interested in the investigation of the differences or, less frequently, similarities between two or more

languages, with reference to some aspect of discourse. The interest centres on the language itself, with the linguistic analysis targeting morphological, syntactical and lexico-grammatical aspects. For instance, Pontrandolfo (2018) studies sentence adverbs in Italian, English and Spanish judicial discourse, taking as the source of data the multilingual corpus COSPE (Pontrandolfo 2016). The hypothesis is that, even though judges are expected to be impartial, their opinion can be traced in adverbs because these are the pragmatic vehicles used to express their stance. The researcher investigates the differences in adverbial use and the discussion of the results focuses on the morphosyntactic level. Similar studies at the linguistic analysis level are Blanco's (2016) investigation of evaluation in English and Spanish newspaper opinion discourse, De Cesare, Albom, *et al*. (2020) on adverbs in Italian, German, Dutch, French and Spanish news media, and Johansson and Rawoens (2019) on impersonal passives in Swedish and Dutch. In many of these studies, the focus has been primarily on differences, and this has been especially so in the case of comparisons oriented towards translation and interpreting where those differences have practical implications. For instance, Bodarenko (2018) investigates a reciprocal English and Russian parallel corpus of dialogue-based fiction (two sub-corpora sections: Russian to English and English to Russian) and examined the differences, and, in light of these differences, the relationships between these two languages in the translation process.

2.2.2. Comparative cultural keyword or discourse keyword studies

The second category is that of comparative cultural keyword or discourse keyword studies which is sufficiently distinctive to merit its own entry. In these studies, the assumption is that because discourse keywords are "semantic nodes in discourses, they allow conclusions about the discourses in which they occur" (Schröter, Veniard, *et al*. 2019: 15). In these studies, by adopting a lexicological approach, the focus of the analysis is concerned with the investigation of the discourse usage and functions of a word or a set of words and the subsequent comparison across cultures and languages. Several papers coming out of the *Discourse Keywords of Migration* project illustrate this. Schröter, Veniard, *et al*. (2019), for example, study the lexical profile of the keywords *multicultural* and *multiculturalism* (multikulturell*, **multiculturel***, *multicultural**) in a comparable multilingual corpus of British, French, German and Italian newspaper articles covering the time span 1998-2012, following a corpus-assisted methodology. Similarly, Taylor (2017) explores the lexical behaviours of *community* and *comunità* in English and Italian in a comparable newspaper corpus, and Schröter and Veniard (2016) compare the use of **intégration** and Integration in French and German public discourses about migration.

Another kind of study in this category would be those such as Jaworska and Leuschner (2018) or Schröter (2018) which address loanwords. Schröter (2018) investigates, at the lexical level, the use of German Nazi vocabulary (for example Anschluss, Judenrein, Blitzkrieg, Lebensraum) in an English webcorpus, and compares the results to those from comparable French and German webcorpora to understand whether the use of German Nazi vocabulary differs between German and the borrowing languages. Jaworska and Leuschner (2018) undertake a study on the discursive transpositions of the Germanism Kulturkampf in German and in two host languages, Polish and English. This study aims to understand whether or not

the meaning of this word is recontextualized to the host languages to fulfil a specific cultural function.

2.2.3. The comparative interest is largely cultural and the research is cross-linguistic out of necessity

The third category is characterized by those studies whose interest is primarily cultural, so language is taken into consideration because of the importance of its role as a rich repository of a culture. For example, Baker and Vessey (2018) conduct a comparison between English and French Islamist extremist texts, using two comparable French and English corpora composed of texts assessed as dangerous and extremist by national authorities. Their aim is to establish how messages in different languages draw upon similar and distinct linguistic strategies related to specific discursive themes and cultural perspectives. Although two languages are used, these are treated as the vehicles for communication, rather than the principal object of study. In another example, Aragrande (2018) analyses how migration discourse is reported in an audio-video corpus with multilingual data from four broadcasting channels, split into two monolingual sub-corpora (*Rai Uno* and *Rainews 24* for Italian and *BBC1* for English) and a bilingual sub-corpus (*Euronews*). Her aim is to analyse different journalistic interventions on migration in the two monolingual sub-corpora and the bilingual corpus in order to compare the results and to define the extent to which the cultural context influences the linguistic representation by comparing the reports collected.

2.2.4. No explicit comparative drive

The fourth type of cross-linguistic CADS is characterized by the absence of comparative aims. The studies in this category generally make use of multilingual data, but this linguistic variety may be related to the same cultural context or dimension. For instance, Freake, Gentil, *et al.* (2011) explore the discursive construction of nationhood and belonging in Quebec in a corpus of public consultation briefs, which were originally submitted in two languages to the Bouchard-Taylor Commission. The corpus was 95% in French and 5% in English and the resulting analysis of both languages was necessary as a matter of completeness. In a different context, Fotopoulos and Kaimaklioti (2016) look at how the Greek, German and British press have reported the refugee crisis. Although they use a multilingual corpus, their interest is to identify a European Media Discourse and the selection of the above-mentioned countries serves to cover both northern and southern Europe rather than to look at the differences among them.

2.3. Corpora in CL-CADS

As can be seen from the discussion above of the different types of studies which analyse discourse across languages using corpus linguistics, the majority of studies do not actually use multilingual corpora. Indeed, in this sense, the area is far more traditional than it might initially appear. The majority of studies make use of multiple monolingual sub-corpora. So, in the case of Del Fante (2018), which investigated the salience of country of origin in representations of immigrants in the British and Italian press before and after the 2016 Brexit referendum, two sets of monolingual

corpora were compiled: two Italian newspapers and two English-language newspapers. The corpora were comparable because they had been collected from similar sources (newspapers) and using similar kinds of search terms to identify articles about migration (though see below for more on the complexities of this). This kind of synchronic comparable monolingual corpora is undoubtedly the most common kind employed in current CL-CADS work. Furthermore, Nardone (2018), who works on German-Italian comparisons, suggests that at least one of the language sub-corpora is frequently English (although this inevitably depends on the publication language).

Another commonly used kind of corpus in cross-linguistic work is the parallel corpus, although here the focus has perhaps less frequently been at the discourse level. In the parallel corpus comparison, again we have multiple monolingual corpora, although in this case they are closely related as one is likely to be the source text and one, or more, the translated text.

Despite important developments in concepts such as *translanguaging*, "the deployment of a speaker's full linguistic repertoire without regard for watchful adherence to the socially and politically defined boundaries of named languages" (Otheguy, García, *et al.* 2015: 281), we far less frequently encounter multilingual corpora in the sense of corpora which contain multiple languages within the same documents/document sets. One notable exception being Freake, Gentil, *et al.* (2011), which is mentioned above. In this, perhaps, CL-CADS shares a blind spot with translation studies more generally, which has similarly remained committed to the concepts of named languages (though see Baynham and Lee 2019).

2.4. Difficulties in designing comparison

As discussed in both Vessey (2013) and Taylor (2014), there are a number of challenges associated with designing comparative studies. These fall into three categories: the design of the comparable corpora which have, to date, formed the bedrock of CL-CADS studies, the design of the analysis, and the interpretation of findings.

2.4.1. Designing the corpora

Regarding the corpora, the first challenge lies in the establishing whether similar text types (for instances, genres) actually exist in the different languages and cultures under examination. To take the example of the press, as this is a frequently studied area in discourse studies, in the UK there has traditionally been a distinction between the tabloid or popular press and the broadsheet or quality press. However, in other countries, that distinction may not exist with more salient binaries being daily vs. weekly, or national vs. regional. In each case, the analyst has to know the function and audience of any given text type in order to assess the potential comparability.

In a discourse-complete corpus, all texts from a given domain and time range are compiled into the corpus. For instance, the *SiBol* corpora (available in *Sketch Engine*) contain all articles published in a range of newspapers in 1993, 2005, 2010 and 2013. Having identified comparable newspapers (mainstream and national), comparable corpora could be designed relatively easily as they would contain all articles published in those time frames. In contrast, in a search-term corpus, which for practical reasons is more common, the second major challenge concerns the

identification of topics. Take the example of the *Discourse Keywords of Migration* project. The aim was to collect a set of comparable corpora of British, French, German and Italian press articles on migration. But how do we decide which terms are equivalent across the four languages when collecting the articles? For instance, in the Italian press discourse, the term *extracomunitario* (literally a person from a non-EU member state) is particularly salient, but has no clear equivalent in British English (the closest in functional terms probably being *illegal immigrant*). Thus, the process of identifying search terms so as to build a set of comparable corpora must also be a phase of the analysis in which functional equivalence is operationalised in a replicable and transparent manner.

2.4.2. Designing the analysis

The second phase of decision-making concerns the analysis and this may take as its starting point either discourse analytic categories (at the *corpus-based* end of the cline in Tognini-Bonelli's 2001 terms) or corpus linguistic tools (more *corpus-driven*). If we consider the basic tools of corpus linguistics, we would anticipate analysis of frequency, keyness or collocation.

The analysis of *frequency*, whether single or multi-word expressions (also known as n-grams) is relatively unproblematic for comparative analyses once the corpora have been designed, although difficulties may arise in the comparison of multiword expressions where the languages being compared are morphologically different.

A very useful measure of frequency for comparison is *keyness*, which tells us the *aboutness* of text and is "a quality words may have in a given text or set of texts, suggesting they are important" and "what the text 'boils down to' […] once we have steamed off the verbiage, the adornment, the blah blah blah" (Scott and Tribble 2006: 55-56). In a monolingual study, *keyness* is usually calculated either by comparing two or more corpora to each other, or to the same reference corpus in order to identify words which are significantly characteristic of the target corpus. In the case of a multilingual study, this becomes impossible and so comparable reference corpora also have to be identified in order to calculate keywords. Furthermore, as Vessey (2013) notes, once obtained, the numerical *keyness* values cannot be reliably compared because they may have been influenced by the reference corpora, rather than being directly attributable to variation in the target corpus. However, in terms of affordances, *keyness* analysis has the strong advantage of providing the analyst with a replicable and comparable starting point for the analysis. It is also corpus-driven in nature and so reduces the researcher's influence on the research direction at this stage.

The other pillar of corpus work is the analysis of *collocation*, those words that go together and have a strong textual relationship. Unlike *keyness*, in order to analyse *collocation*, the analysis has a lexical starting point; the analyst has to decide which words will constitute comparable nodes. This carries with it the same need to check that terms are functionally equivalent and therefore, like identifying search terms for corpus construction, collocation studies require comparative analysis before starting. We discuss a collocation case study in Section 3.1.

So far, we have considered the three classical corpus linguistics starting points. However, a corpus and discourse study may equally start from the discourse tools and then use corpora to find evidence of occurrence. For instance, corpus work could

be employed in identifying whether a particular social actor is likely to be the "doer" or "done-to" in a given corpus. In the case of discourse analytic starting points, the use of categories (such as *individuation* vs. *collectivisation* in the representation of migrants, as applied in Lams 2018) rather than a lexical starting point, facilitates the analytic design.

In Sections 3 and 4, we discuss methods of designing analysis at both the lexical level and at the supra-lexical level.

2.4.3. Interpretation

The last area of challenge regards the interpretation of findings. The primary point to make here is that the analyst must be cautious in attributing any variation identified in cross-linguistic discourse studies to the variables of language or culture when other factors could have intervened.

This might involve over-interpretation of findings as cautioned against in Truan (2019), in which she compares *people*, Volk, and **peuple** in British, German, and French parliamentary debates. The cultural connotations of Volk in Germany mean that the term is greatly underused compared to the other two items and so the analysis needed to include Mensch which "takes on some of the features of *people*, especially the need to speak on people's behalf" (Truan 2019: 224).

In a similar vein, Vessey (2013: 10) raises the problem of over-simplifying the relationship between population and sample (that is, the corpus) and notes that:

> [i]n Canada, where language is used as an identity marker across the population (see, for example, Karim 1993; and Kymlicka 1998: 10), associating an English language corpus with English speakers and a French corpus with French speakers is over-simplistic, since this essentialises and reifies the differences between the two and overlooks the potential role of other groups.

Thus, in any cross-linguistic study it is essential that the researcher be sufficiently familiar with the context of production to enable them to restrain interpretation and expand the analysis where required.

3. Analysis at the lexical level

At the most fundamental level, a corpus-assisted discourse study may involve comparison at the lexical level. That is to say, we take two or more comparable terms and examine the lexical profiles in the different languages under study. This may be the end point of the analysis, or, as discussed above, it may be a preparatory stage which is required to identify search terms for compiling the corpora, or comparable terms to use for analysis of collocation. In many ways, this is the most "traditional" starting point and the closest to the lexicographical origins of corpus linguistics. As Nordrum comments regarding contrastive linguistics:

> […] there is a focus on studying issues from the perspective of form to function rather than from function to form […] one reason for this limitation in focus, of course, is the strong influence of corpus linguistics methodology on contrastive linguistics […] corpus methodology simply lends itself to contrastive studies below clause level. To extend the perspective to the discourse level, manual analysis is still required. (Nordrum 2015: 239)

3.1. Principles for comparison at the lexical level

In preparing the analysis, the researcher needs to identify equivalents in the different languages. As readers within the field of translation and interpreting studies will know only too well, this is never a simple task, nor one that can be solved with recourse to a bilingual dictionary alone.

According to Hoey (2005: 13), who derived his theory of language through corpus analysis, "every word is primed for use in discourse as a result of the cumulative effects of an individual's encounters with the word." Within a given domain/genre, he argues more specifically that:

- Every word is primed to occur with other words; these are collocates;
- Every word is primed to occur with particular sets; these are its semantic associations;
- Every word is primed to occur in association with particular pragmatic functions; these are the pragmatic associations;
- Every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations;
- Every word is primed for use in one or more grammatical roles; these are its grammatical categories;
- Every word is primed to participate in, or avoid, particular types of cohesive relation in a discourse; these are its textual collocations;
- Every word is primed to occur in particular semantic relations in the discourse; these are its textual semantic associations;
- Every word is primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations. (Hoey 2005: 13)

Collectively, these *primings* form what it means to know to a word and thus may act as a guide for the analyst seeking equivalents for a cross-linguistic study. Furthermore, there are even more basic questions of use. If we take the example again of comparing English and Italian migration discourses, we might note that *immigrante* is often offered as a translation of *immigrant*, but, as discussed in Taylor (2014), the difference in frequency of use is so great that they cannot be considered equivalent.

In practice, what this means is that a range of sources need to be drawn upon to identify the nodes for a comparison at the lexical level; from dictionaries, to metacommentary (where available), to frequency comparisons and collocation analysis.

3.2. Case-study: identification of discourse keywords of migration

The *Discourse Keywords of Migration* project originated as a research network with the aim of creating an online European dictionary of migration discourse keywords. Discourse keywords are understood as salient lexical items that occur frequently in certain discourse contexts. The notion draws on traditions of cultural keywords and conceptual history with "emphasis on those cultural keywords which have sociopolitical significance in a particular period" (Jeffries and Walker 2018: 4), alongside the corpus linguistics notions of *keyness* (as mentioned above).

One of the challenges in this type of analysis is finding a way of defining and operationalising the concept such that it can be repeated across different linguistic and cultural contexts. As part of this project (first discussed in Schröter and Veniard 2016), discourse keywords were operationalised as lexical items that:

> a) occur frequently especially in periods of salience of the discourse it belongs to (for example, austerity in the discourse about the financial crisis since 2008); b) function as lexical nodes in discourse, focal points of discourse-determined semantic accumulation that upon deeper analysis unravel a part of the history and ideology of the underlying discourse; c) are usually part of an ensemble of other lexical items that feature prominently in the same discourse context; "there is not just one keyword that labels the related discourse (e.g. migration), but typically there are a number of such nodes, and the discourse key words inhabiting them (e.g. multicultural society) often represent certain points of view (e.g. fortress Europe) or are established as a (counter) reaction to others (e.g. illegalised immigrants vs. illegal immigrants)"; d) refer to issues that are controversially debated; thereby, the labelling can be controversial (such as illegal(ised) immigrants), or the issue at hand can be controversially evaluated without questioning the adequacy of the label (namely, whether or not a multicultural society is something positive). (Schröter and Veniard 2016: 4)

Taylor (2017) explored the discourse keywords *community* and *comunità* in the British and Italian press. These are particularly interesting terms to investigate because there are contrasting reports of their evaluative potential, or in Hoey's (2005) terms the pragmatic meanings for which they are primed. Works on both English (Williams 1983) and Italian (Gallissot, Kilani, *et al.* 2007) terms from the cultural keywords tradition emphasise the favourable meanings and uses. In contrast, work combining (critical) discourse analysis and corpus linguistics, Baker, Gabrielatos, *et al*. (2013) and Taylor (2009) both noted the pattern in which *community* describes those who are "other" to the speaker. This apparently contrasts with Freake, Gentil, *et al*. (2011) who find that *community* is used to self-describe English-speaking groups in Quebec, but they present themselves as minorities within a majority French-affiliated society.

Community and *comunità* were determined to be discourse keywords for migration in both contexts on the basis that:

1) They display peaks of frequency which correspond to politically and socially relevant moments;
2) They both occur as collocates of key migration-related terms (for example, *immigrant* and *immigrate*);
3) They were the subject of metadiscussion in the press articles indicating some controversy around the use of these terms.

Thus, from a preliminary analysis, it was clear that the two terms fulfil three of the four criteria for a discourse keyword (the second is more challenging to operationalise). On the basis of this, it would then be possible to extend the analysis to other languages to see if a word such as **communauté** also functions as a discourse keyword for migration in French, or other languages. The subsequent analysis showed that both terms, as used in migration discourses in the mainstream press, were predominantly used to describe "others" and out-groups. Furthermore, the term was avoided when describing an in-group, which in this context is more likely to be conceptualised as *society*.

To sum up, the principle challenge of starting the analysis at a lexical level is that of equivalence while one of the key affordances to this kind of approach is that it can give us a "key" to the discourses that are under examination and it is fully replicable across additional languages.

4. Analysis above the lexical level

In the previous section, we discussed ways in which an analysis can develop from a lexical starting point. However, in many cases we may wish to carry out the analysis at a higher level of abstraction so in this section we survey some of the methods which have been used to achieve this. This is not intended as a complete guide, but draws on case studies which we have worked on. The intention is that they will provide scope for discussing some of the affordances and challenges involved in designing cross-linguistic analyses, and, longer-term, act as a starting point for the collation of methods. A key reflective part of research is pulling together work to get an overview of the tools we have available to help our mutual endeavour of better understanding language use.

4.1. Requirements to allow for comparison between languages

The first basic requirement for analysis in any cross-linguistic, corpus-assisted discourse study is that there be a *systematic* and *replicable* frame so that it can be repeated on new language corpora with confidence. The second constraint is that any analysis of discourse needs to be done with an understanding of what discourse is. That is, one must resist the temptation to "count what is easy to count" thanks to the corpus tools.

4.2. Comparison of semantic groupings

One method of abstracting out above the lexical level is through some form of semantic categorisation. In this case the starting point is still evidently lexical, but we are not comparing the different language corpora directly at that level. We move up a level to make the comparison in terms of emerging categories or *semantic associations* in Hoey's (2005) terms.

4.2.1. Comparison of key semantic categories

Corpus linguistics provides some tools for accomplishing this level of categorisation with semantic tagging. For instance, Song, Chin-Chuan, *et al.* (2019) use *WMatrix* (Rayson 2008) to compare US, UK, and Chinese media coverage of "China's rise" from 2009 to 2017. Using a form of keyness analysis, they find that four semantic categories in the Chinese corpus outnumber those in the Anglo-American corpus: abstract terms; relationship and social actions; science, technology and culture; and communication. In contrast, the categories which are stronger in the Anglo-American corpus are money and commerce, and politics and warfare. These differences in keyness of semantic categories are interpreted in terms of a Chinese preference for framing the rise in terms of soft-power while in the English-language press that was analysed, it is a more overt form of economic and military might. The system on which this analysis was based, USAS (*UCREL Semantic Analysis System*), has been developed to automatically tag corpora in Finnish, Russian, Italian, Chinese, Portuguese, Spanish, Dutch, Czech, Urdu, Malay, Arabic, and Welsh (Piao, Rayson, *et al.* 2016).

An alternative method for comparison at this level would be to calculate keywords for each comparable corpus (calculated based on a comparable reference

corpora) and subsequently manually group the keywords into semantic sets. This could be done with reference to existing semantic sets, or, if made fully transparent, it may be the analyst who derives the categories. This is the method used in the next case study.

4.2.2. Comparison of semantic associations in collocates

The *semantic associations* and *pragmatic associations* which Hoey (2005) lists as part of the primings of a word also provide a way to compare word usage at a more abstract level. In the *Discourse Keywords of Migration* project, these were used to compare collocates of discourse keywords across languages. Collocates are significant in investigating discourse because "collocation analysis offers a productive means for understanding ideology, as lexical co-occurrence may shed new light on complex webs of identities, discourses and social representations in a community" (Bogetić 2013: 334).

Schroter, Veniard, *et al.* (2019) report the findings of a comparison of the discourse keywords German multikulturell*, French **multiculturel***, English *multicultural*, and Italian *multicultural** based on their collocates. The collocates in each case were calculated based on using comparable corpora and the same software (*CQPWeb*), the same measure of collocation (*loglikelihood*), the same span for collocates (5L/R) and the same cut-off for collocation (200). The authors independently categorised the collocates in each language based on examining the term's usage in expanded concordance lines. The authors subsequently compared their classifications and collectively identified a shared template. This is an essential stage in any multilingual project in order to avoid the data, which is in the lingua franca (in this case, as so often, English), from becoming the default base for analysis, subsequently imposing a single view on the classifications emerging from other language corpora.

An extract from the table presenting the collocates is shown in Table 1. As can be seen from Table 1, even without translation, the frame-based presentation and abstraction to the level of semantic sets allows us to see where there is similarity across the language sub-corpora (presence of collocates in the same set) and difference (sparsely populated or empty cells). Schröter and Storjohann (2015: 49) argue that identifying slots "that remain empty in a template matrix frame, capturing what could be (but is not) said about the financial crisis" is informative for understanding the discourse "since what is phenomenologically absent can still be epistemologically relevant."

The analysis coming out of this initial classification showed that while the keyword was highly frequent in German immigration debates, it had relatively few collocates compared to the other three languages, indicating a less coalesced usage. Furthermore, as can be seen above, in the German discourse, multicultural was not systematically associated with processes, again pointing towards a different conceptualisation. Other patterns that were identified include a more negative association evident in the French, German and Italian press (in the table below, the negative evaluating terms are marked in bold). Each of these points served as the starting point for further analysis and the study showed how the formally related words were all discourse keywords of migration, but with greatly differing degrees of saliency and with clearly distinct usages, particularly in the case of German.

TABLE 1

Extract from the table presented in Schröter, Veniard, *et al.* (2019: 25-27)

Semantic Category	Related Collocates (English)	Related Collocates (French)	Related Collocates (German)	Related Collocates (Italian)
Related Concepts	multi-ethnic, multiracial, melting + pot, mix, nation, backgrounds, community, tolerance, identity, diversity, communities, immigration, national	métissée, multi-ethnique, **mosaïque,** intégration, identité	Zusammen-leben, Mitein-ander	multietnica, multireli-giosa, multirazziale, multireli-gioso, integrazione, multietniche, multietnico, interetnico, interculturale, **razzismo**
People	Muslims, immigrant, population, black, white, group, immigrants, minister	partisans, immigrants, gens, enfants, On	Wir	Merkel, autori, direttore, cittadino, popolo, abitanti, Leader
Actions	creating, become, believe	devenue, devenir, **limites**, mutation, créer, attendre, veut, tente, développement, **cause** (in remettre en cause, criticize), doit, peut, faut, va		**confrontano**, viviamo, gestito, diventando, diventata, riconosce, riservata, rendere, costruire, attraverso, essere, diventare, dobbiamo

4.2.3. Affordances and challenges of comparison of semantic categories

The process of semantic categorisation of collocates allows for a data-driven entry point into the data. The focus on collocates is a corpus linguistics entry point, but one which relies on semantic and discourse-grounded understandings of how language works. The abstraction to semantic categories for both collocates and keywords allows for comparison above the lexical level and the matrix presentation allows readers to see for themselves what the raw findings look like, even without understanding all languages studied.

In terms of challenges, the micro-level processes are fundamental to obtaining meaningful findings. First, the categorisation must be based on contextualised occurrences; that is the analyst needs to go beyond considering the isolated word to examining how it is used in the specific context, in this case of press discourses on migration. If we recall, Hoey's hypotheses about primings are bound by *genre*. Second, it is unavoidable that the categorisations will be subjective. Each researcher may potentially see different groups. Subjectivity is an inevitable part of the work we do and so the crucial element lies in recognising this and accounting for it in our research. The full list of categorised items should be presented to allow others to see the process. And, in the case of multi-authored contributions, categorisation of each language set should be carried out independently to avoid the imposition of categories that might suffocate emerging patterns found only in one of the language corpora. We would argue this is particularly important in studies that include English

given the ease with which this may become the "default" structuring language for analysis. Lastly, given the subjective nature of the classification, it is important to "look the other way" and try to counter a corroboration drive (Taylor and Marchi 2018), that is our natural tendency to look for further confirmation of what we have already found. For instance, when it emerged from the analysis of *community* and *comunità* that these terms predominately described out-groups, Taylor (2017) looked for evidence that this might not be the case by searching for references to these discourse keywords in co-occurrence with *British* or *Italian** (although, even in these cases, the writer was not talking about their own group, but minority groups such as *British* + [other nationality] + *community* or diasporic groups such as *comunità italiana* + [foreign nationality/city]).

4.3. Comparison of discourse frames: moral panics

While the previous section discussed corpus linguistic starting points for abstracting out above the lexical level, in the next two sections we consider how discourse analytic tools may form the basis for designing a comparative analysis.

4.3.1. Case-study: moral panic frames

The case study presented here is that of the moral panic frame which is applied to the analysis of immigration discourses in the British and Italian press in Taylor (2014). The academic theorisation of moral panic begins with Cohen who describes how:

> [s]ocieties appear to be subject, every now and then, to periods of moral panic. A condition, episode, person or group of persons emerges to become defined as a threat to societal values and interests; its nature is presented in a stylized and stereotypical fashion by the mass media; the moral barricades are manned by editors, bishops, politicians and other right-thinking people; socially accredited experts pronounce their diagnoses and solutions; ways of coping are evolved or (more often) resorted to; the condition then disappears, submerges or deteriorates and becomes more visible. (Cohen 1972: 9)

In the analysis of migration discourse, which concerns the authors of this paper, this appears highly applicable to contemporary debates and indeed has been used in Maneri (2001) to investigate Italian migration discourse. Drawing on previous research and his own qualitative observations, McEnery (2005) develops and tests Cohen's set of roles in moral panic discourse and adds a seventh category relating to the language of moral panic discourse. These roles/categories are:

1) "Object of offence," that which is identified as problematic;
2) "Scapegoat," that which is the cause of, or which propagates the cause of, offence;
3) "Moral entrepreneur," the person/group campaigning against the object of offence;
4) "Consequence," the negative results which it is claimed will follow from a failure to eliminate the object of offence;
5) "Corrective action," the actions to be taken to eliminate the object of offence;
6) "Desired outcome," the positive results which will follow from the elimination of the object of offence;
7) "Rhetoric," register marked by a strong reliance on evaluative lexis that is polar and extreme. (adapted from McEnery 2005: 6-7)

Taylor (2014) investigates the nationalities which are found to be foregrounded in the British and Italian press (this included people from Afghanistan, Iraq and Romania in the British press and people from Libya, Somalia, and Eritrea in the Italian press). The analysis was done first by categorising the transitivity roles which members of these groups are assigned when they are mentioned (for example, whether they are represented as "doing" or "receiving" actions). In the next stage, this information was used to verify whether each of the seven moral panic roles was present in the press reporting. For instance, in the case of migrants from Afghanistan in the UK tabloid newspapers, they were repeatedly associated with an "object of offence" which was presenting a threat to resources in the UK. The consequences were depicted as negative for UK citizens through illustrative and personalised stories of people waiting for housing, and there was consistent intensification of the rhetoric with highly negative collocates (*illegal*, *arrested*) typifying their representation and dehumanising metaphors. However, there was no moral entrepreneur or corrective action presented. This was typical of the reporting of all foregrounded groups in both the English and Italian language data: there was little evidence of fully-fledged moral panic at that time. This was perhaps counter to what one might expect, given the highly negative reporting which was present, but serves to illustrate how having a replicable frame can allow us to empirically test intuitions.

4.3.2. Affordances and challenges

The advantage to taking a discourse frame as the basis for comparison is that the analysis is theoretically meaningful. In the case of the moral panic frame, this is a model which has been developed for discourse studies. It allows the analyst to abstract above the purely lexical level to examine patterns in different sets of data, irrespective of the language. In the case of a fully iterated frame, like the moral panic model, the roles can be used to create a matrix grid to readily identify absences and presences. Lastly, as seen above, the highly systematic nature of the frame means that the analyst is pushed to consider all facets and thus it counters the corroboration drive and allows for unexpected results to emerge.

The challenges involve, first, identifying a discourse frame which is relevant and applicable. Second, identifying systematic and replicable methods for filling the slots in the frames. For instance, in the model above, the "object of offence" slot was considered filled if the migrants were consistently depicted as actors of actions that have a detrimental effect (from the speaker's viewpoint).

4.4. Comparison of rhetorical features: metaphor

Another discourse analytic starting point is through the comparison of use of rhetorical features. This might include the identification and analysis of pronouns to investigate deixis, or, in the case study below, the use of metaphor.

4.4.1. Case-study: metaphors of migration in the press in the USA and Italy

Metaphor use is particularly salient in investigations of discourse because it is a pervasive concept which involves a group of linguistic, cognitive, pragmatic, rhetoric properties that are present at the same time in varying degrees (Charteris-Black 2004; Patterson 2018) and, so, is of great interest due to "its evaluative potential, whereby

selected favourable or unfavourable elements of the source are projected onto the target" (Partington, Duguid, *et al.* 2013: 131). Metaphors play at a conceptual and discourse level, letting the researcher outflank the problems of lexical comparability discussed in the previous sections. In this way, a cross-linguistic approach takes advantage in terms of comparability. As claimed by Deignan and Potter (2004: 1232):

> [b]ecause conceptual metaphor theory claims to describe central processes and structures of human thought, it is not language-specific and should have explanatory power for languages other than English; it is therefore of potential use in cross-linguistic research.

There are several studies which adopt a cross-linguistic critical metaphor approach. Schmidt (2002) conducted a corpus-based study on the presence of a set of conceptual metaphors in three corpora of business reports in Finnish, Finland-Swedish, and German. Similarly, Deignan and Potter (2004) examined two English and Italian corpora to explain the behaviour of *body* metaphors (for instance, *break the heart*) in two different linguistic and cultural contexts. Dervinytė (2009) investigated conceptual emigration and immigration metaphors and their linguistic manifestations in a corpus of British and Lithuanian press articles. More recently, Bisiada (2018) compared the use of the *homework* metaphor (*done their homework*; *done their job*) in two German and English corpora. What comes through from the literature is that, although every language and every culture has its own metaphors, we can hypothesise the presence of an underpinning structure: an object or process is interpreted through a mapping from a source domain to a target domain. This structure is the common ground for a cross-linguistic comparison and such features are perfectly suited to contrastive study.

Most work concerning the metaphorical representation of migrants (Santa Ana 1999; El Rafaie 2001; Semino 2008; Dervinytė 2009; Hart 2011, 2014; among others) have shown how newspaper discourse and public discourse in general have presented a predominantly negative representation of *immigrants/immigration* through the selection of specific conventional metaphors. For instance, they have been represented as *animals*, depicting them as non-humans, not just non-citizens. Other studies have reported the use of natural disaster metaphors, in particular *water* metaphors realised with expressions like *flood of* or *tides of*. The only exception is represented by Salahshour (2016) who shows that in New Zealand economic discourse immigrants are positive represented as a *force* which is *gained* by the country, and Taylor (2018) who shows that in the 2018 British parliamentary debates the migrants belonging to the Windrush Generation are positively represented as *builders* within a *country as house* frame. With an understanding of how metaphors are used differently, it would be easier to shed light on the cognitive process through which metaphors work and on how the context influences some metaphor and defines their use (Lakoff 1987; Goatly 1997).

Del Fante (forthcoming) compares the representation of migrants in newspapers in Italy and the USA over the period 2000-2005. Two comparable corpora were created using the *LexisNexis* database and the search terms *migrant** and *immigrant** for English and *migrant** and *immigrat** for Italian. As the author was interested in conventional metaphorical expression related to movement and quantity, all occurrences of the following patterns were retrieved and manually searched for potential metaphor: ? + *di immigrat**/ *di migrant**/ *de* immigrat**/*de* migrant** for Italian; ? + *of*

immigrant/migrant** for English. Considering that conventionality is strongly related to frequency (Charteris-Black 2004), in order to get only the most relevant results, *LogLikelihood* was used as the statistical measure for both language sub-corpora.

The results showed the presence of the same conceptual metaphors in both countries under analysis. Examination of the top 100 collocates for both corpora showed the presence of the conceptual metaphor *migrations are liquid* (*flusso*, *flussi*, *ondata*, *ondate*; *influx*, *wave*, *waves*, *flow*, *tide*, *flood*) which has been extensively discussed in previous literature (El Rafaie 2001; Charteris-Black 2006; Salahshour 2016; Taylor 2018; among others). Immigrants are represented, in both newspapers, as a large and uncontrolled body and the implicature is that such movement should be controlled. As Charteris-Black (2006) claims, there is a relation between liquid metaphors and control: a liquid moves differently and may be difficult to control. A lack of control over movements is also a lack of control over changes in societies and, therefore, over societies. For this reason, the metaphor often argues that migration has to be controlled.

The second grouping which emerged in the top 100 collocates was *migrants are a multitude*, for which the definition as metaphor is arguable. This included expressions such as *migliaia*, *quote*, *centinaia*, *milione*, *decine*, *quota*, *centinaio*, *cento*; *millions*, *hundreds*, *numbers*, *tens*, *dozens*, *percentage*. In this case, we do not have metaphor in the traditional sense. Expressions like *hundreds of* or *millions of* are not deviant from the normal use of language and their function is to express the quantity, so these do not satisfy classic requirements for being metaphorical. However, we can hypothesise that these findings would discursively reinforce the presence of other metaphors like *liquid* metaphors and *invader* metaphors. These lexical items indicating an indefinite large quantity show a *semantic association* with migration-related terms, establishing a pattern. This pattern contributes to the representation of migration as an indefinite, numerous and problematic group of people (Baker 2006: 79), which is the basic implicature for other metaphors. For example, as for the *liquid* metaphor, liquids including water are dangerous if there is a large uncontrolled quantity.

This brief case-study shows how metaphor analysis can reveal the presence of rhetorical features of language even from a cross-linguistic and cross-cultural perspective. Despite the difference between the two countries under analysis, it showed strong similarities in the metaphorical representation of migrants. In addition to the presence of the same conceptual metaphors, the same linguistic metaphors were found to represent the same topic positing the existence of a kind of western migration discourse in the press, strongly characterized by a privileged perspective for the writers.

4.4.2. Affordances and challenges

The analysis of metaphor gives the researcher a means for abstracting out above the lexical level so that findings can be compared across different language corpora. Crucially, we have a theoretical frame which explains the relevance of metaphor for discourse construction and so there is a conceptually sound base for the choice of features. We know from previous research that metaphors work to structure discourse and have a substantial evaluative role which can operate below the level of conscious word choice.

However, there are challenges associated with cross-linguistic analysis. First, the researchers need to decide how they will use the corpora to identify the metaphors, for instance, searching by references to the source or target of the conceptual metaphor. Second, they will have to decide what will be included as a metaphor, something which is always less clear-cut than it might appear. Third, we have to be very cautious in the interpretation of metaphor; it is not the case that the same metaphor will always carry the same evaluation. To take the example of *water* metaphors, in a study of historical discourse, Taylor (forthcoming) finds that *water* metaphors were used to favourably describe migrants when they were viewed by the speaker as an economic resource to be exploited (the metaphor running along the lines of "migrants are the water that powers the mill").

5. Conclusions

Corpus linguistics has proved to be an immensely valuable partner for (critical) discourse analysis, as evidenced in the multiplicity of work in this area, the emergence of a new journal (*Journal of Corpus and Discourse Studies*) and a dedicated conference series (*Corpora and Discourse International Conference*). As such, we can expect it to constitute an important *new context* for translation and interpretation-oriented discourse studies, the theme of this special issue.

We have shown here that there are a number of ways in which we can combine corpus linguistics and discourse analysis in order to analyse corpora in different languages, both at the lexical level, and above the lexical level, and we hope that these methods may be of interest and relevance to researchers working on discourse analysis in relation to translation (and interpreting, although we acknowledge that the tendency of corpus work to focus on written texts makes this crossover more challenging).

As we stated at the beginning, we have not attempted to provide an exhaustive taxonomy of approaches, but we do intend for this paper to serve as a starting point for building such a resource. It is the nature of scientific progress that researchers in the field need to know what is already available and what has already been tested in order to develop our fields—and collaboration is the only way for this to occur.

The main point that we hope to have made is that *reflexivity*, the self-awareness of what we are doing, why, and what impact that choice may have, needs to lie at the centre of our cross-linguistic corpus-assisted discourse research. Reflexivity in research is, of course, a general principle, but it is especially pertinent in cross-linguistic studies because of the multiplicity of variables that we have to handle. It is also a tradition that, as Baker (2018) suggests, is particularly well embedded within corpus linguistics. As corpus linguistics is relatively new compared to other linguistic methodologies, those who employ it often feel the need to justify and explain the underpinning assumptions and procedures. Thus,

> […] it has perhaps resulted in those working within corpus linguistics being particularly reflective on their method and the claims that are made about it. Or it may be the distinct combination of qualitative and quantitative techniques that has made us more apt to think about method *per se*. (Baker 2018: 291)

NOTES

* For the purposes of Italian evaluations, the authors note that Dario del Fante was the lead author of Sections 2.2 and 4.4.

1. The studies discussed in the following sections are only those which use corpora in their comparative work. Pre-corpus studies are of great relevance, but outside the scope of this paper.

REFERENCES

Ancarno, Clyde (2018): Interdisciplinary approaches in corpus linguistics and CADS. *In*: Charlotte Taylor and Anna Marchi, eds. *Corpus Approaches to Discourse: A Critical Review*. London/New York: Routledge, 130-156.

Aragrande, Gaia (2018): Migrants in Translation: A Corpus-Based Approach to the Representation of Migrants by Four News Broadcasting Channels. *In*: Irene Ranzato and Serenella Zanotti, eds. *Linguistic and Cultural Representation in Audiovisual Translation*. London/New York: Routledge, 121-146.

Baker, Paul (2018): Conclusion: Reflecting on reflective research. *In*: Charlotte Taylor and Anna Marchi, eds. *Corpus Approaches to Discourse: A Critical Review*. London/New York: Routledge, 281-292.

Baker, Paul and Vessey, Rachelle (2018): A corpus-driven comparison of English and French Islamist extremist texts. *International Journal of Corpus Linguistics*. 23(3):255-278.

Baker, Paul, Gabrielatos, Costas, and McEnery, Tony (2013): *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.

Baker, Paul (2006): *Using Corpora in Discourse Analysis*. London: A&C Black.

Baynham, Mike and Lee, Tong King (2019): *Translation and Translanguaging*. London/New York: Routledge.

Bisiada, Mario (2018): A cross-linguistic analysis of the 'homework' metaphor in German and English political discourse. *Discourse & Society*. 29(6):609-628.

Blanco, Maria Perez (2016): 'An important issue': a corpus-based contrastive study of evaluation along the parameter of relevance in English and Spanish newspaper opinion discourse. *RAEL: Revista Electrónica de Lingüística Aplicada*. 15(1):36-52.

Bogetić, Ksenia (2013): Normal straight gays: Lexical collocations and ideologies of masculinity in personal ads of Serbian gay teenagers. *Gender and Language*. 7(3):333-367.

Bondarenko, Antonina (2018): A corpus-based contrastive study of verbless sentences: quantitative and qualitative perspectives. *Studia Neophilologica*. 1:1-24.

Charteris-Black, Jonathan (2004): *Corpus Approaches to Critical Metaphor Analysis*. Basingstoke: Palgrave Macmillan.

Charteris-Black, Jonathan (2006): Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse & Society*. 17(5):563-581.

De Cesare, Anna Maria, Albom, Ana, Cimmino, Doriana, *et al.* (2020): Domain adverbials in the news: A corpus-based contrastive study of English, German, French, Italian and Spanish. *Languages in Contrast*. 20(1):31-57.

Deignan, Alice and Potter, Liz (2004): A corpus study of metaphors and metonyms in English and Italian. *Journal of Pragmatics*. 36(7):1231-1252.

Del Fante, Dario (2018): Where Are They From? A Corpus-Assisted Study of the Geographical Representation of Migrants in the UK and Italian Quality Press in 2016. *I-LanD Journal – Identity, Language and Diversity*. 1:28-52.

Del Fante, Dario (forthcoming): *The metaphorical representation of migration across time and cultures: the cases of USA and Italy*. Doctoral thesis, unpublished. Padua/Brighton: University of Padua/University of Sussex.

Dervinytė, Inga (2009): Conceptual emigration and immigration metaphors in the language of the press: A contrastive analysis. *Kalbų studijos*. 14:49-55.

El Refaie, Elisabeth (2001): Metaphors we discriminate by: Naturalized themes in Austrian newspaper articles about asylum seekers. *Journal of Sociolinguistics*. 5(3):352-371.

Fairclough, Norman (1989): *Language and Power*. London: Longman.

Fotopoulos, Stergios and Kaimaklioti, Margarita (2016): Media discourse on the refugee crisis: on what have the Greek, German and British press focused? *European View*. 15(2):265-279.

Freake, Rachelle, Gentil, Guillaume, and Sheyholislami, Jaffer (2011): A bilingual corpus-assisted discourse study of the construction of nationhood and belonging in Quebec. *Discourse & Society*. 22(1):21-47.

Gallissot, René, Kilani, Mondher, and Rivera, Annamaria (2007): *L'imbroglio etnico in quattordici parole-chiave* [Ethnic cheating in fourteen keywords]. Bari: Edizioni Dedalo.

Goatly, Andrew (1997): *The Language of Metaphors*. London/New York: Routledge.

Hart, Christopher (2011): Moving beyond metaphor in the cognitive linguistic approach to CDA. *Critical Discourse Studies in Context and Cognition*. 43:171-192.

Hart, Christopher (2014): *Discourse, Grammar and Ideology: Functional and Cognitive Perspectives*. London: Bloomsbury.

Hoey, Michael (2005): *Lexical Priming: A New Theory of Words and Language*. London/New York: Routledge.

Jaworska, Sylvia and Leuschner, Torsten (2018): Crossing languages – crossing discourses. A corpus-assisted study of Kulturkampf in German, Polish and English. *Pragmatics and Society*. 9(1):117-147.

Jeffries, Lesley and Walker, Brian (2018): *Keywords in the Press: The New Labour Years*. London: Bloomsbury.

Johansson, Annika and Rawoens, Gudrun (2019): A corpus-based contrastive study of impersonal passives in Swedish and Dutch. *Languages in Contrast*. 19(1):2-26.

Karim, Karim H. (1993): Construction, deconstructions and reconstructions: competing Canadian discourses on ethnocultural terminology. *Canadian Journal of Communication*. 18(2). Consulted on 1 March 2020 <https://www.cjc-online.ca/index.php/journal/article/view/744>.

Kymlicka, Will (1998): *Finding our Way: Rethinking Ethnocultural Relations in Canada*. Oxford: Oxford University Press.

Lakoff, George (1987): *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago.

Lams, Lut (2018): Discursive constructions of the summer 2015 refugee crisis: A comparative analysis of French, Dutch, Belgian francophone and British centre-of-right press narratives. *Journal of Applied Journalism & Media Studies*. 7(1):103-127

Maneri, Marcello (2001): Il panico morale come dispositivo di trasformazione dell'insicurezza [Moral panic as a device for transforming insecurity]. *Rassegna Italiana di Sociologia*. 1:5-40.

Mautner, Gerlinde (2015): Checks and balances: How corpus linguistics can contribute to CDA. *In*: Ruth Wodak and Michael Meyer, eds. *Methods of Critical Discourse Studies*. London: Sage, 154-179.

McEnery, Tony (2005): *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. London/New York: Routledge.

Nardone, Chiara (2018): 'Women and Work': A cross-linguistic corpus-assisted discourse study in German and in Italian. *Critical Approaches to Discourse Analysis Across Disciplines*. 10(1):167-186.

Nordrum, Lene (2015): Review of contrastive discourse analysis: functional and corpus perspectives. *Linguistics and the Human Sciences*. 10(3):327-332.

Otheguy, Riccardo, García, Ofelia, and Reid, Wallis (2015): Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review*. 6(3):281-307.

PARTINGTON, Alan, DUGUID, Alison, and TAYLOR, Charlotte (2013): *Patterns and Meanings in Discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam/Philadelphia: John Benjamins.

PATTERSON, Katie (2018): *Understanding Metaphor through Corpora: A Case Study of Metaphors in Nineteenth Century Writing*. London/New York: Routledge.

PIAO, Scott, RAYSON, Paul, ARCHER, Dawn, *et al.* (2016): Lexical Coverage Evaluation of Large-scale Multilingual Semantic Lexicons for Twelve Languages. *In*: Nicoletta CALZOLARI, Khalid CHOUKRI, Thierry DECLERCK, *et al.*, eds. *Proceedings of the 10th edition of the Language Resources and Evaluation Conference* (LREC 2016, Portorož, 23-28 May 2016). Paris: European Language Resources Association. Consulted on 22 June 2019, <http://www.lrec-conf.org/proceedings/lrec2016/index.html>.

PONTRANDOLFO, Gianluca (2016): *Fraseología y lenguaje judicial. Las sentencias penales desde una perspectiva contrastiva*. Roma: Aracne.

PONTRANDOLFO, Gianluca (2018): Sensibly and appropriately the judge considered. A corpus-based study of sentence adverbs in judicial language. *Linguistik online*. 92(5):215-234.

RAYMOND, William (1983): *Keywords: A Vocabulary of Culture and Society*. London/New York: Routledge.

RAYSON, Paul (2008): From key words to key semantic domains. *International Journal of Corpus Linguistics*. 13(4):519-549.

SALAHSHOUR, Neda (2016): Liquid metaphors as positive evaluations: A corpus-assisted discourse analysis of the representation of migrants in a daily New Zealand newspaper. *Discourse, Context & Media*. 13:73-81.

SANTA ANA, Otto (1999): Like an animal I was treated: Anti-immigrant metaphor in US public discourse. *Discourse & Society*. 10(2):191-224.

SCHMIDT, Christopher (2002): Metaphor and cognition: a cross-cultural study of indigenous and universal constructs in stock exchange reports. *Journal of Intercultural Communication*. 5:17p. Consulted on 30 June 2019, <http://www.immi.se/intercultural/abstracts/schmidt-abs.htm>.

SCHRÖTER, Melani (2018): How words behave in other languages: the use of German Nazi vocabulary in English. *Pragmatics and Society*. 9(1):93-118.

SCHRÖTER, Melani and VENIARD, Marie (2016): Contrastive analysis of keywords in discourses: *intégration* and *integration* in French and German discourses about migration. *International Journal of Language and Culture*. 3(1):1-33.

SCHRÖTER, Melani and STORJOHANN, Petra (2015): Patterns of discourse semantics: A corpus-assisted study of financial crisis in British newspaper discourse in 2009. *Pragmatics and Society*. 6(1):43-66.

SCHRÖTER, Melani, VENIARD, Marie, TAYLOR, Charlotte, *et al.* (2019): A comparative analysis of the keyword multicultural(ism) in French, British, German and Italian migration discourse. *In*: Andreas MUSOLFF and Lorella VIOLA, eds. *Migration and Media: Discourses about Identities In Crisis*. Amsterdam/Philadelphia: John Benjamins, 13-44.

SCOTT, Michael and TRIBBLE, Chris (2006): *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam/Philadelphie: John Benjamins.

SEMINO, Elena (2008): *Metaphor in Discourse*. Cambridge: Cambridge University Press.

SONG, Yunya, CHIN-CHUAN, Lee, and ZEPING, Huang (2019): The news prism of nationalism versus globalism: How does the US, UK and Chinese elite press cover 'China's rise'? *Journalism*. 20p. Consulted on 30 June 2019, <https://journals.sagepub.com/doi/abs/10.1177/1464884919847143>.

TAYLOR, Charlotte and MARCHI, Anna, eds. (2018): *Corpus Approaches to Discourse: A Critical Review*. London/New York: Routledge.

TAYLOR, Charlotte (2009): The representation of immigrants in the Italian press. *CirCap Occasional Papers*. 21:40p. Consulted on 22 June 2019, <https://www.circap.org/occasional-papers.html>.

Taylor, Charlotte (2014): Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. *International Journal of Corpus Linguistics*. 19(3):368-400.

Taylor, Charlotte (2017): Togetherness or othering? 'Community' and 'comunità' in the UK and Italian press. *In*: Jan Chovanec and Katarzyna Molek-Kozakowska, eds. *Representing the Other in European Media Discourses*. Amsterdam/Philadelphia: John Benjamins, 55-80.

Taylor, Charlotte (2018): Representing the Windrush generation: metaphor in discourses then and now. *Critical Discourse Studies*. 17(1):1-21. Consulted on June 20, 2019, <https://www.tandfonline.com/doi/ref/10.1080/17405904.2018.1554535>.

Taylor, Charlotte (forthcoming): *Migration Discourses and Memory*. London/New York: Routledge.

Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins.

Truan, Naomi (2019): The discursive construction of the people in European political discourse: Semantics and pragmatics of a contested concept in German, French, and British parliamentary debates. *In*: Jan Zienkowski and Ruth Breeze, eds. *Imagining the Peoples of Europe. Populist discourses across the political spectrum*. Amsterdam/Philadelphia: John Benjamins, 201-228.

Vessey, Rachelle (2013): Challenges in cross-linguistic corpus-assisted discourse studies. *Corpora*. 8(1):1-26.