TRANSLATION-INDUCED LANGUAGE CHANGE IN THE FIELD OF MIGRATION: A MULTILINGUAL CORPUS ANALYSIS OF EU LEGAL TEXTS AND PRESS ARTICLES

By

EDWARD CLAY

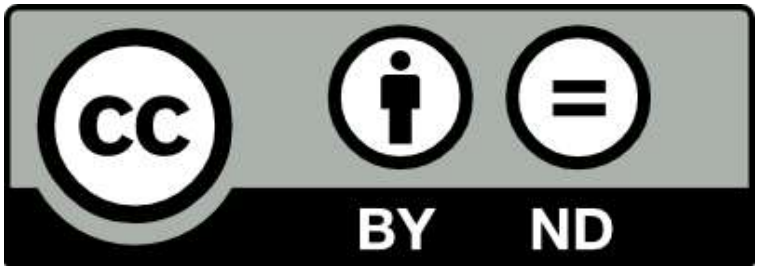
A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Modern Languages
School of Languages, Cultures, Art History and Music
College of Arts and Law
University of Birmingham
November 2023

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive
e-theses repository



This unpublished thesis/dissertation is under a Creative Commons Attribution- NoDerivatives 4.0 International (CC BY-ND 4.0) licence.

You are free to:

Share — copy and redistribute the material in any medium or format for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NoDerivatives — If you remix, transform, or build upon the material, you may not distribute the modified material.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

Unless otherwise stated, any material in this thesis/dissertation that is cited to a third party source is not included in the terms of this licence. Please refer to the original source(s) for licencing conditions of any quotes, images or other material cited to a third party.

ABSTRACT

The terminology used in both legal and non-legal contexts to discuss issues related to migration is particularly critical given the consequences that the use of such language can have on people's lived experiences and the ways they are perceived by others. It is therefore essential to understand how this terminology is evolving and the way in which it interacts across different languages. While the potential of translation to act as a propagating factor in language change across a range of language combinations has been demonstrated by previous research (e.g. Bisiada 2016, Malamatidou 2011, Amouzadeh & House 2010), this is the first sustained study investigating translation-induced language change in the field of migration terminology. This research aims to examine changes in migration terminology in English, French and Italian EU legal texts and press articles over a pre-defined time frame (1992-1998 & 2015-2018). The research focuses on the analysis of two purpose-built multilingual corpora, the *EUMigrLaw Corpus* and the *MigrationInTheNews Corpus*, created using the online corpus analysis software Sketch Engine. The *EUMigrLaw Corpus* is a multilingual parallel corpus made up of a range of binding and non-binding legislative texts produced by the EU, and the *MigrationInTheNews Corpus* is a comparable corpus containing a range of press articles from newspapers from the UK, France and Italy on the subject of migration. The research methodology, broken down into three sections of results and analysis, combines a range of both existing and innovative quantitative and qualitative corpus methods to detect and measure intralingual and interlingual variation and change in both corpora and identify potential sites of translation-induced language change in migration terminology. The key findings from this research reveal i) persistent widespread terminological variation across all three languages in EU legal texts in

the field of migration and areas of terminological convergence among the languages being studied, suggesting the occurrence of translation-induced language change; ii) intralingual terminological variation in press articles in each language, with significant differences between the terminological trends identified in each language; iii) isolated examples of contact-induced changes originating in EU texts, before becoming de-terminologised and appearing in press articles. This study therefore provides a rigorous methodological approach to examining different types of terminological variation and change in migration terminology over time, revealing the potential and limitations of translation-induced language change in this specific context. This thesis consequently explores the significant legal and linguistic implications of its findings and considers subsequent avenues for future research.

ACKNOWLEDGEMENTS

I would first and foremost like to thank my family who, in countless different ways, have helped and supported me through the PhD process. My parents and my partner, Tonderai, have been a constant source of encouragement and belief from the very beginning, and my children, R&R, who arrived midway through, have brought me joy, inspiration and perspective in even the most difficult of moments. Without them, this thesis would have remained unwritten.

I would also like to thank my wonderful supervisors, Dr Sofia Malamatidou and Professor Aleksandra Čavoški, for their unwavering support and honesty, as well as their warmth and kindness. They made immeasurable contributions throughout, helping me set sail on my journey, enabling me to navigate choppy waters and seeing me back to safe harbour. I consider myself immensely fortunate to have benefited from their vast experience, boundless knowledge and thoughtful guidance.

Finally, I would like to express my gratitude to the Wolfson Foundation for providing the financial support which made this research possible.

Contents

1. Introduction 1
1.1 Overview 1
1.2 Objectives 3
1.3 Methods 8
1.4 Outline of thesis 9
2. Overview of existing literature 12
2.1 Translation-Induced Language Change 12
2.2 Legal Translation in the EU Context 21
2.3 The Language of Migration in Law and the Print Media 35
2.4 Terminological Variation 46
3. General Methodology 54
3.1 Corpus-Based Approach 54
3.2 *EUMigrLaw Corpus* 55
3.2.1 The Development of EU Immigration and Asylum Law 59
3.2.2 Legal Acts in the European Union 64
3.2.3 Building the *EUMigrLaw Corpus* 68
3.3 *MigrationInTheNews Corpus* 70
3.3.1 Building the *MigrationInTheNews Corpus* 72
3.4 Step-by-step methodological approach 78
4. Stage 1: *EUMigrLaw Corpus* Analysis 81
4.1 Methodology 81
4.2 Quantitative analysis 92
4.2.1 Intralingual terminological variation 92
4.2.2 Interlingual terminological variation 101
4.3 Qualitative analysis 108
4.3.1 General observations 108
4.3.2 Influence of English terminology 112
4.3.3 The case of IRREGULAR MIGRANT 119
4.4 Summary of results 123
4.5 Discussion 126
5. Stage 2: *MigrationInTheNews Corpus* analysis 135
5.1 Methodology 135
5.2 Quantitative results 138
5.3 Terminological trends in conceptual categories 145
5.3.1 ASYLUM SEEKER and REFUGEE 145

5.3.2 MIGRANT 152
5.3.3 IRREGULAR MIGRANT 161
5.3.4 MIGRANT WORKER 171
5.3.5 THIRD-COUNTRY NATIONAL 177
5.3.6 UNION CITIZEN 183
5.4 Discussion 188
6. Stage 3: Terminology comparison between the *EUMigrLaw Corpus* and the *MigrationInTheNews Corpus* 193
6.1 Methodology 193
6.2 Quantitative results on proportion of concepts per subcorpora 197
6.3 Quantitative results on terminological convergence/divergence 200
6.4 Terminological trends 207
6.4.1 THIRD-COUNTRY NATIONAL 207
6.4.2 IRREGULAR MIGRANT 210
6.4.3 UNION CITIZEN 213
6.4.4 ASYLUM-SEEKER and REFUGEE 216
6.4.5 MIGRANT and MIGRANT WORKER 219
6.5 Discussion 223
7. Conclusion 232
References 243
Legal Sources 272
Appendix 1.1: English to French *EUMigrLaw Corpus* 1992-1998 277
Appendix 1.2: French to Italian *EUMigrLaw Corpus* 1992-1998 281
Appendix 1.3: Italian to English *EUMigrLaw Corpus* 1992-1998 285
Appendix 1.4: English to French *EUMigrLaw Corpus* 2015-2018 290
Appendix 1.5: French to Italian *EUMigrLaw Corpus* 2015-2018 295
Appendix 1.6: Italian to English *EUMigrLaw Corpus* 2015-2018 300
Appendix 2: Full results tables for Stage 2: *MigrationInTheNews Corpus* analysis 305
Appendix 3: Stage 3: Proportion of concepts per subcorpora 313

List of figures

Figure 1: The Pillar Structure of the EU after the Maastricht Treaty (ILSP and MCRC, 2000) 60
Figure 2: Terms used in English in 1992-1998 subcorpus for concept UNION CITIZEN 96
Figure 3: Terms used in French in 1992-1998 subcorpus for concept UNION CITIZEN 96
Figure 4: Terms used in Italian in 1992-1998 subcorpus for concept UNION CITIZEN 97
Figure 5: Terms used in English in 2015-2108 subcorpus for concept UNION CITIZEN 98
Figure 6: Terms used in French in 2015-2018 subcorpus for concept UNION CITIZEN 99
Figure 7: Terms used in Italian in 2015-2018 subcorpus for concept UNION CITIZEN 99
Figure 8: Distribution of mean IVI values by language pair and time period 104
Figure 9: percentage of terms showing a statistically significant change in frequency 142
Figure 10: Percentage change in relative frequency of terms in conceptual categories ASYLUM SEEKER and REFUGEE for English news articles 145
Figure 11: Percentage change in relative frequency of terms in conceptual categories ASYLUM SEEKER and REFUGEE for French news articles 148
Figure 12: Percentage change in relative frequency of terms in conceptual categories ASYLUM SEEKER and REFUGEE for Italian news articles 149
Figure 13: Percentage change in relative frequency of terms in conceptual category MIGRANT for English news articles 152
Figure 14: Percentage change in relative frequency of terms in conceptual category MIGRANT for French news articles 156
Figure 15: Percentage change in relative frequency of terms in conceptual category MIGRANT for Italian news articles 158
Figure 16: Percentage change in relative frequency of terms in conceptual category IRREGULAR MIGRANT for English news articles 161
Figure 17: Percentage change in relative frequency of terms in conceptual category IRREGULAR MIGRANT for French news articles 164
Figure 18: Percentage change in relative frequency of terms in conceptual category IRREGULAR MIGRANT for Italian news articles 167
Figure 19: Percentage change in relative frequency of terms in conceptual category MIGRANT WORKER for English news articles 171
Figure 20: Percentage change in relative frequency of terms in conceptual category MIGRANT WORKER for French news articles 173
Figure 21: Percentage change in relative frequency of terms in conceptual category THIRD-COUNTRY NATIONAL for English news articles 177
Figure 22: Percentage change in relative frequency of terms in conceptual category THIRD-COUNTRY NATIONAL for French and Italian news articles 180
Figure 23: Percentage change in relative frequency of terms in conceptual category UNION CITIZEN for English news articles 183
Figure 24: Percentage change in relative frequency of terms in conceptual category UNION CITIZEN for French and Italian news articles 185
Figure 25: Each concept in each subcorpora as an overall proportion of English terms 198
Figure 26: Each concept in each subcorpora as an overall proportion of French terms 199
Figure 27: Graph showing each concept in each subcorpora as an overall proportion of Italian terms 200
Figure 28: Results of indicators 1, 2 and 3 for English for both time periods 201
Figure 29: Results of indicators 1, 2 and 3 for Italian for both time periods 202
Figure 30: Results of indicators 1, 2 and 3 for French for both time periods 203

Figure 31: Change in prevalence of predominant term in conceptual category THIRD-COUNTRY NATIONAL in the EUMigrLaw Corpus 208
Figure 32: Proportion of terms in conceptual category IRREGULAR MIGRANT containing adjectives 'irregular', 'irrégulier' and 'irregolare' in EU texts and news articles 213
Figure 33: Percentage differences between frequencies of terms across the EUMigrLaw Corpus and the MigrationInTheNews Corpus for both time periods for concepts ASYLUM-SEEKER and REFUGEE 218

List of Tables

Table 1: Number of tokens in EUMigrLaw Corpus....57
Table 2: Legally binding and non-legally binding documents included in the EUMigrLaw Corpus....65
Table 3: Number of tokens in MigrationInTheNews Corpus....71
Table 4: News sources in the MigrationInTheNews Corpus by language....73
Table 5: Terms in the conceptual category UNION CITIZEN for English (1992-1998)....83
Table 6: Conceptual category for MIGRANT WORKER (English, 2015-2018)....85
Table 7: Examples of term pairs allocated a terminological distance value of 0....86
Table 8: Examples of term pairs allocated a terminological distance value of 0.5....87
Table 9: Examples of term pairs allocated a terminological distance value of 1....88
Table 10: Example of the calculation of IVI for UNACCOMPANIED MINOR (IT-EN 2015-2018)....90
Table 11: Conceptual categories in the earlier and later time periods....92
Table 12: Number of terms in each subcorpus by language....94
Table 13: Term pairs for DISPLACED PERSON in earlier subcorpus....102
Table 14: Mean IVI values....105
Table 15: p-values derived from Fisher's Exact test for each language pair....106
Table 16: THIRD-COUNTRY NATIONAL conceptual category: Italian....138
Table 17: Number of terms in MigrationInTheNews Corpus by language and time period....139
Table 18: Overall proportions of concepts....143
Table 19: Comparative results table for conceptual category MIGRANT WORKER (Eng 1992-1998)....195
Table 20: Example of category with same predominant term....196
Table 21: Predominant terms in English for UNION CITIZEN conceptual category....215
Table 22: Predominant terms in French for conceptual categories MIGRANT and MIGRANT WORKER....220

1. Introduction

1.1 Overview

This thesis presents the first sustained study of translation-induced language change on terminology in the field of migration. The theory of translation-induced language change considers translation to be a specific kind of language contact and argues that under certain conditions it can propagate changes in the lexical, grammatical and communicative properties of the receiving language (House, 2006). In order to investigate the applicability of this theory to migration-related terminology, this study adopts a range of methodological approaches using two purpose-built multilingual corpora made up of English, French and Italian texts to observe the terminological dynamics occurring within them, explore the reasons for the phenomena identified and discuss their implications. The corpora used in this study are made up of EU legal texts (*EUMigrLaw Corpus*) and press articles (*MigrationInTheNews Corpus*), which are subject to extensive quantitative and qualitative analysis to identify changes in migration terminology both within and between the languages in question. The changes examined in this study are both synchronic (occurring at the same time) and diachronic (occurring over a given time period), providing an insight into the complex dynamics shaping the terminology used in this field in the specialist genre of EU legislation and the non-specialist genre of press articles.

Terminology in the field of migration has long been a central point of focus in legal circles and an ongoing bone of contention in the press. Legal scholars have frequently debated term usage in this domain (e.g. Bauder, 2014; Kostas, 2017; Pérez-Paredes, Jiménez & Hernández, 2017) while national and supranational institutions have often felt the need to debate and clarify the terminology that should

be used to refer to specific concepts. The UN Special Rapporteur on the Human Rights of Migrants, for instance, underlined that 'incorrect terminology contributes to negative discourses on migration, reinforces negative stereotypes against migrants, and legitimates a discourse of the criminalisation of migration which in turn contributes to further alienation, marginalisation, discrimination and violence against migrants' (Crépeau, 2013). In the media, such topics are also hotly debated, leading to headlines such as 'Why are white people expats when the rest of us are immigrants?' (Koutonin, 2015) in the UK newspaper *The Guardian*, '*Migranti, rifugiati, sfollati, profughi: mettiamo ordine nel linguaggio politicamente corretto*' [Migrants, refugees, displaced persons, asylum seekers: making sense of politically correct language] (dal Monte, 2018) in the Italian newspaper *Il Fatto Quotidiano*, and '*Immigration. Petites phrases, gros préjugés'* [Immigration. Small phrases, large prejudices] (Urbach, 2023) in the French newspaper, *l'Humanité*. These articles, and many more, discuss the frequent misuse of terms in the press, exploring the lexical nuances involved in the complex domain of immigration and asylum, arguing that the incorrect use of terminology in such a sensitive area could result in the stigmatisation of individuals and even put them in positions of danger. These examples serve to demonstrate the level of interest and scrutiny in the legal sphere and in the media regarding the language used to talk about migration and the increasing recognition that terminology can have a genuine impact on people's lives, influencing everything from their legal status to the way they are perceived by the general public, making this an appealing and thought-provoking lens through which to investigate translation-induced language change.

This study adopts a three-stage approach, starting with an analysis of the terminological dynamics in EU legal texts. This initial stage examines the extent of

terminological variation both within the same language (intralingual variation) and across different languages (interlingual variation). It explores how key concepts in the field of migration have been expressed by different term variants in a range of legislative texts. For the purposes of this research, terms are broadly defined as lexical units that are used to express a concept in a specific domain (see chapter 2.4 for more details). This first stage then investigates the extent to which the terminology used in the three languages in question has converged in the field of EU migration law to provide an insight into the potential occurrence of translation-induced language change in this environment. The following stage examines migration terminology used in press articles in the UK, France and Italy. This stage explores how key migration concepts are expressed in the languages in question in a non-specialist context and how these term variants change over time. The final analysis stage uses a range of metrics to conduct a comparison of the changes detected in migration terminology in the two genres. The analysis carried out at this final stage sheds light on the degree of influence of translation-induced changes identified in EU legislative texts on terminology used in press articles.

1.2 Objectives

Previous research has pointed to the prevalence of variation in terminology in EU legal texts, whereby different term variants are used to refer to the same concept (Biel, Biernacka & Jopek-Bosiacka, 2018; Peruzzo, 2013; Prieto Ramos & Morales Moreno, 2019), while other studies have highlighted the capacity of translation to play a role in propagating language change in certain circumstances (Malamatidou, 2016; Amouzadeh & House, 2010; Bisiada, 2016). It has also been argued that the

EU language system constitutes a propitious site for translation-induced language change to take place (Clay, forthcoming). Indeed, among the relevant factors listed by Kranich, Becher and Höder in their typology of translation-induced language change (2011) are ‘intensity of contact’, ‘length of contact’, ‘prestige of SL’ and ‘typological proximity’ (pp.18), all of which can be applied to the language contact scenario being studied in the EU context, as explored in greater detail in chapter 2. This thesis therefore brings these two concepts of terminological variation and translation-induced language change in EU texts together as part of an empirical study for the first time to explore the hypothesis that there is a link between the terminological variation detected in these texts and the translation system used to produce them. One key aim is therefore firstly to determine whether terminological variation can be detected in the terminology used in migration-related legislation produced by the EU institutions, and then to establish the extent to which translation may be involved in the changes identified.

In addition, this thesis seeks to fill another gap in the current literature by exploring whether translation-induced changes in one genre can cross into other genres. In this case, it aims to determine whether such changes in legal language can influence terminology in news articles. Indeed, previous studies on translation-induced language change have been limited to examining change within a specific genre, while this study attempts to detect the potential reach of this phenomenon beyond the boundaries of the genre in which it originates. It is reasonable to hypothesise that those changes arising in EU legal language may influence terminology used in non-specialist press articles given the prestige of the EU institutions, their increasing influence over migration-related matters and the volume of new legislation produced on the matter that is likely to be of general interest to readers. Indeed, supranational

organisations such as the EU have both the technical and linguistic resources to become ‘terminological benchmarks’ in their areas of competence with the capacity to have an impact on lexical choices in other settings (Prieto Ramos & Morales Moreno, 2019).

From a linguistic perspective, this research was therefore motivated by the need to gain an understanding using empirical, data-driven methods as to whether the theory of translation-induced language change can be applied in a number of novel ways: to EU legal texts, in relation to migration terminology and in the case of changes crossing genre boundaries. This is significant since it provides an original insight into the impact that translation (and translators) can have on the evolution of terminology in various ways that are underrepresented in previous research and how this terminology interacts across different languages and genres.

Another principal goal of this study is to gain a better understanding of the forces shaping migration-related terminology, both in the specialist legal domain and in general, non-specialist usage. The critical importance of terminology in these domains is undeniable. In the legal sphere, terms used to categorise people at defined points in the migration process (asylum seeker, refugee, returnee, irregular migrant, etc.) can have a pivotal effect on the rights to which these people are entitled and the obligations that states have towards them. Legal language should not be assumed to always be neutral and should be evaluated critically through rigorous methods such as those implemented in this thesis.

EU law was identified as a particularly relevant subject of study in this context since it sets the economic, social and political parameters governing very large numbers of people living in and entering Europe and also because it is subject to translation

activities on an unprecedently vast scale. On a broader level, the legal terminology used in migration legislation also helps to shape the wider discourse around the issue of migration and how it is perceived by citizens and the media. This thesis seeks to make a significant contribution to this field because, while there is widespread recognition of the importance of this issue, there has as yet been no systematic study into the factors shaping and contributing to the evolution of this terminology in EU law.

Moreover, given the unique linguistic environment of the EU institutions, scholars have hypothesized that language itself plays a critical role in integration and policymaking (De Swaan, 2001). Much more than a by-product of its composition, the EU's multilingualism 'is directly linked to the political nature of the union' (Felici, 2010, pp. 95), and is an integral part of how the system functions. Indeed, it is considered by some researchers that the delicate balance achieved by the EU language system has contributed to the relative success of its legislative creation process and depoliticized its approach to policymaking (Ringe, 2022; Mamadouh, 1999). In light of this, an investigation into the language of EU legal texts must also inevitably have implications for the political and decision-making structures of the institutions. Therefore, aside from the purely linguistic findings, the results of this study can also provide insights into the development and evolution of the EU in more general terms.

In addition, it has been argued that public discourse in the media can in turn influence migration policy and legal discourse in the field (e.g Berkeley, Khan & Ambikaipaker, 2006; Somerville, 2007). This research will also provide an insight on whether there is evidence from terminological research of the corpora that indicates

an impact from the print media on legal language, and more broadly the extent to which these two spheres interact with regard to terminology use over time.

The language of the print media is also particularly influential in forming public opinion, defining salient perspectives and establishing interpretational lenses on unfolding events (Gruessing and Boomgarden, 2017), which also makes it a particularly interesting focus of study in this regard. Much research has been carried out into the language used to talk about migration in the print media, the consequences that such language can have on people's lived experiences and how they are perceived by others (e.g. Blinder & Allen, 2016; Khosravinik, 2009; Berry, Garcia-Blanco & Moore, 2015), however no previous study has yet examined the potential impact of EU legal terminology on the language of the press in different countries. This thesis therefore also contributes to the field of migration studies by providing an in-depth, empirical account of the changes in how people at various points in the migration process are referred to in English, French and Italian in both EU law and the press.

Bearing in mind the various objectives set out here and the motivations for them, the overarching aim of this thesis is therefore to answer the following research questions:

i) Can terminological variation be detected in EU migration law terminology, whether diachronic, synchronic, interlingual or intralingual in nature?

ii) Is such variation partly attributable to translation-induced language change?

iii) Is there evidence of such translation-induced changes crossing genre boundaries and becoming established in press articles?

1.3 Methods

This research uses a range of quantitative and qualitative methods from corpus linguistics in order to provide a comprehensive analysis of the phenomenon under examination. The two multilingual corpora used in this study (*EUMigrLaw Corpus* and *MigrationInTheNews Corpus*) were purpose-built using the online corpus management software Sketch Engine[1] (Kilgariff et al., 2014), consisting of approximately 2.5 million words each, and cover two distinct time periods (1992-1998 & 2015-2018) in order to track terminological change over time (see chapter 3 for more detail). The corpus-based approach adopted for this study was chosen to support a rigorously empirical, data-driven investigation into language variation and use, in order to provide findings with greater generalizability and validity than would otherwise be feasible with manual analysis (Biber, 2015). A number of novel methodological approaches are adopted throughout this thesis, which use quantifiable metrics to provide measurable results regarding terminological convergence or divergence across different languages. The quantitative results derived from the various stages of corpus analysis are also subjected to statistical significance testing to strengthen the validity of the findings. While highly informative and useful at providing a good general overview of patterns and dynamics, these quantitative analyses are not sufficient by themselves to draw reliable conclusions,

[1] www.sketchengine.eu

hence the use of complementary qualitative analyses throughout in order to provide a more fine-grained picture of the phenomena being studied. The methodology used in this thesis is also innovative insofar as it involves the concurrent analysis of three different languages: English, French and Italian. Previous studies examining translation-induced language change have mainly focused on language pairs (principally English and another language). This novel approach makes it possible to measure degrees of influence across different languages to examine whether languages behave in the same way under the influence of translation and also offers a more comprehensive picture of the development of terminology in this area than would otherwise be possible from an investigation involving only two languages.

1.4 Outline of thesis

Chapter 2 of this thesis provides a comprehensive overview of the current literature relevant to the topics explored in this thesis. It indicates where this current study is situated in relation to research previously conducted and sets out the theoretical underpinnings on which this thesis is based. This chapter is split into four sections, which provide a review of the relevant literature in relation to i) translation-induced language change, ii) the language of EU law, iii) the language of migration in law and the press, and iv) terminological variation.

Chapter 3 sets out the general methodological approach adopted for this research study. It explains in detail the content and creation process of both corpora used in the analysis, including some necessary background on the development of migration law in the EU and the press landscape in the UK, France and Italy. The specific

methodological approaches used in each of the three separate analyses are described in detail in chapters 4, 5 and 6.

Chapter 4 presents the first stage of the three-part results and analysis process. This part involves the analysis of the *EUMigrLaw Corpus*, evaluating synchronic and diachronic intralingual and interlingual terminological variation in EU migration law. It also determines the extent of terminological convergence or divergence among the three languages being studied in order to draw conclusions regarding the occurrence of translation-induced language change in this setting.

Chapter 5 sets out the findings of the analysis of the *MigrationInTheNews Corpus*. This second stage of analysis identifies the extent of synchronic and diachronic terminological variation within each language in press articles in English, French and Italian on the subject of migration. This chapter also examines the similarities and differences between the terminological patterns and trends emerging from each language.

The third stage of analysis is presented in chapter 6, which provides a synchronic and diachronic comparison between the migration terminology in the *EUMigrLaw Corpus* and the *MigrationInTheNews Corpus* in order to identify similarities and differences in the terminological trends observed in the text types contained in the corpora. The results of the quantitative and qualitative analysis set out in this chapter indicate the extent to which translation-induced language change has the capacity to cross genre boundaries.

Chapter 7 offers a conclusion of the research conducted in this thesis, providing a summary of the key findings and a discussion of the main legal and linguistic

implications of these results. It also discusses the limitations of the study and considers subsequent avenues for future research.

2. Overview of existing literature

Prior to addressing the original contributions set out in this thesis, it is important to situate it in the wider research context. This chapter sets out a comprehensive overview of previous research relevant to the study presented in this thesis. It is made up for four sections, providing an overview of i) translation-induced language change, ii) legal translation in the EU context, iii) the language of migration in law and the print media, and iv) terminological variation.

2.1 Translation-Induced Language Change

Languages are constantly evolving, and the changes that occur are the result of innumerable forces acting upon them. Some changes in a language can be attributed, at least in part, to contact with other languages. We can broadly define contact-induced language change for the purposes of this thesis as ‘any linguistic change that would have been less likely to occur outside a particular contact situation’ (Thomason, 2001, pp. 62). Studies into language contact have traditionally tended to focus on direct spoken interaction between two or more language communities (e.g. Thomason & Kaufman, 1988; Heine & Kuteva, 2005), which unarguably constitutes a fruitful area for gaining insights into contact-induced language change. However, until relatively recently, the role of written discourse in contact-induced language change had been largely overlooked. A number of translation studies scholars conducting research in this area have argued that translation could be described as constituting a special kind of language contact environment (e.g. Baumgarten, House & Probst, 2004; Kranich, 2014; Bisiada 2016; Malamatidou 2016), with the translator effectively moving back and forth between the

source and target texts. These studies have shown that, much like in other language contact scenarios, under certain circumstances, this kind of contact can lead to changes in the languages involved over time.

In many modern-day translation environments, the goal of the translator is to maintain the ‘function’ the original text has ‘in its discourse world’ (House, 2006, pp. 25-6). This approach is what House refers to as ‘covert translation’ (1997, pp. 69), whereby the translator must take account of the underlying cultural differences so that the translated text resembles an original source text in its own right. This involves passing the source text through a ‘cultural filter’ to ensure that the end product conforms to the expectations of the target reader. However, if, for whatever reason, this cultural filter is not applied, then specific features of the source language may remain in the target text. This phenomenon has been described as the source text ‘shining through’ (Teich, 2003) onto the target text, with traces of the source language still detectable in the translation. This phenomenon can be thought of as the source language exerting an influence over the target language by adapting its normal linguistic features, conventions and patterns. From this, innovations of use may become commonplace in target language texts as a result of this source language influence. These may be limited solely to translated texts in the target language but may occasionally also be incorporated into the repertoire of the target language, at least within specific domains (such as the legal domain) (Kranich, Becher & Höder, 2011; Malamatidou, 2016).

Much of the research in this area has examined the position of English as a lingua franca, and the effects of its predominance over other languages. A series of studies was carried out as part of the ‘Covert Translation’ project at the University of Hamburg (e.g. House, 2006; Becher, House & Kranich, 2009; Amouzadeh & House,

2010) with a view to investigating the potential effect of translation on language change, specifically focusing on English as a global lingua franca. This research project consists of separate studies united by the underlying hypothesis that the dominance of English in various domains, such as popular science, business and economics, has an impact on text production in other languages, initially in translations from English and ultimately in original text production in those languages too. The linguistic point of focus for the studies in this project are communicative norms specific to the languages under examination. The analysis of English originals and their German translations in Baumgarten, House and Probst (2004) is essentially inconclusive as to whether the high rate of translations from English in economic, scientific and computer-related texts has led to a discontinuation of cultural filtering, but points to the possibility that German textual norms may be at a stage of transition. It is argued that adaptations with regard to the 'linguistic expression of stance' and 'addressee involvement' could indicate the initial signs of a more widespread change taking place among German communicative norms as a result of contact with English through translation (pp. 105).

Further research as part of the Covert Translation project provided more evidence of contact-induced language change through translation, including studies by House (2006; 2011) which conclude that discourse features, such as 'for example' and 'for instance' and degrees of subjectivity, in German translations and original texts change under the influence of English, adapting to English conventions. To demonstrate that such changes are not limited to the English-German translation pairing, the project also examines the effect of English as a global lingua franca in translation on Persian (Amouzadeh & House, 2010), showing that the same phenomenon of translation-induced language change can still occur between more

typologically distant languages. Indeed, this study shows that intensive contact with English through translation increases the use of the previously infrequent passive voice in Persian, indicating the occurrence of translation-induced language change in the area of syntax. The effect of this language contact was not to introduce an entirely new syntactic structure into Persian, but to use existing grammatical structures in the target language in a novel way, with greatly increased frequency. Indeed, it is also posited by Hansen-Schirra (2011, pp. 155) that the phenomenon of change in this context most often occurs through 'atypical frequency distributions of typical structures' which are promoted by the process of translation.

The studies described above laid a strong foundation for the study of contact-induced language change through translation, and this thesis will adopt a similar underlying hypothesis on the potential of translation to play a propagating role in language change. However, rather than focusing on a specific genre, such as economic or business texts, it will examine the linguistic features associated with a specific topic – migration – across different genres, both legal and journalistic. In this way, it aims to gain a broader understanding of how changes to the terminology of a specific topic may be propagated by translation and the factors which promote or inhibit such shifts.

Subsequent studies were then undertaken to examine this phenomenon further, adopting rigorous, corpus-based approaches. Using a diachronic corpus study, Bisiada (2016) further supported earlier findings that concessive conjunctions in German are increasingly used in sentence-initial position, which mirrors the conventions in English. Since the results of this study show that this shift occurred first in translations and then in non-translations, it provides evidence that this change may have been propagated by translation rather than simply reflecting changes

occurring in the development of the German language more generally. In a similar study, Pang & Wang (2020) point to the change in the use of adversative conjunctions in Chinese as indicating the influence of English in translation.

Beyond communicative features, previous research has also examined a range of lexical and morphosyntactic features of different languages. Malamatidou (2016) demonstrates that translation of popular science articles from English to Greek has played a role in increasing the use of the passive voice in the popular science target texts in Greek. It is suggested that the increase in the use of this non-typical syntactic structure in Greek popular science articles is a result of large-scale translation from English in this genre, and that in time, the frequency of the passive voice in non-translated Greek popular science articles may well in turn increase and converge with the patterns seen in the translated texts. These findings echo the conclusions of Amouzadeh & House (2010) in supporting Steiner's (2008, pp. 321-2) argument that changes initially emerge as atypical frequencies of existing linguistic features for a significant period of time, before the 'pressure builds' sufficiently for the language change to fully take hold.

A range of other studies have focused specifically on the languages under investigation in this thesis, as well as English, namely French and Italian. Laviosa (2006) argues that the lexical impact of English through translation in certain varieties of Italian is extensive, while Musacchio (2005, pp. 94) claims that 'English-Italian translation is slowly but subtly influencing the conventions governing some genres', pointing to syntactic and lexical changes, such as, for instance, a shift in conventions in Italian economic texts moving away from long sentences with complex linguistic connections towards conventions which are more common in English economics texts. Research into news translations from English into Italian

(McLaughlin, 2013) argues that syntactic shifts in Italian may be propagated by translation from English, with the preposing of usually postposed adjectives seemingly encouraged by the English model. A similar study into translation from English to French (McLaughlin, 2010) points to the influence of contact on the syntax of the translations, suggesting that this may result in syntactic borrowing in contemporary French. While these kinds of synchronic study of translation-induced language change can locate potential areas of influence, they are unable to track changes over time. Indeed, with regard to the influence of English, Bennett (2011, pp. 197-8) questions whether translation is even required for such changes to occur, or whether English discourse features might simply be entering other languages 'through a process of imitation' due to the general prevalence of English in all domains. To overcome such uncertainties, this thesis adopts a diachronic approach as used in various other corpus-based studies in the field (e.g. Malamatidou, 2016; Bisiada, 2016) in order to demonstrate empirically whether a change appears in translated texts first or simply mimics shifts that have already occurred in the language in general use.

Studies by researchers in other language pairings have shown the potential source-text influence on the use of the speaker-inclusive passive in Estonian and German translations into Finnish (Kolehmainen & Riionheimo, 2016) and the influence of popular science translations from English into Chinese (Liao, 2010). In her study, Liao argues that the interactive features in both translated and non-translated Chinese popular science texts have departed from the Chinese norms, and that the translations can be seen as 'the agent that effects the changes' (pp. 139) rather than simply reflecting them, with active adaptation of conventions by writers of Chinese non-translations in order to conform to the interactive features of the Chinese

translations of English. This may be explained in part by the relative prestige enjoyed by English, particularly in the domain of popular science, which makes it more likely that certain features may be copied in other languages, especially when a genre is in its infancy, as in this case of popular science in Taiwan. This collection of studies in a range of language pairs constitutes a solid body of evidence that translation can play a role in propagating contact-induced language change. Consequently, like these studies before it, this thesis aims to respond to the call for 'an array of connected studies that build on each other' based on the observation that this is 'the only way to address the difficult question of whether language contact through translation influences language change' (Bisiada, 2013, pp. 2). As mentioned above, this thesis is also innovative in that it examines changes across three languages rather than a language pair, which allows for direct comparison between how the different languages behave in this specific contact scenario over time.

Despite the similar findings arising from this range of studies, there is not unanimity regarding the extent of the role played by translation in encouraging contact-induced language change to occur. For instance, in examining the influence of English on Italian translations of technical documents, focusing on the transfer of lexical items, Bernardini & Ferraresi (2011) find that translators are in fact less likely to use anglicisms than technical writers of Italian non-translated comparable texts, arguing that in this case, Toury's 'law of growing standardisation' predominates over the 'law of interference' (Toury, 1995). In this scenario, it is hypothesized that the translators prefer the more typical options presented by the target language system and are less comfortable using anglicisms than technical writers of comparable texts in Italian. Indeed, in many cases, it can be very difficult to identify the cause(s) of a given change. Neumann (2011) suggests using additional methods such as eye-tracking to

ascertain the extent to which translators are actually influenced by the source text but admits that nothing can provide conclusive causal evidence or account for all the possible factors which may lead to the adoption of changes. As a result of this reality, it is generally accepted by language contact scholars that a variety of both internal and external factors are involved in language change and that it is important to resist 'the urge to offer a single simple explanation for all types of structural change' (Thomason, 2020, pp. 33). Indeed, there is a grey area between the specific effects of translation and those of language contact more generally, making the investigation into these phenomena particularly challenging. Therefore, when researching contact-induced language change through translation, rather than ruling out other potential triggers, this study will aim to provide as much evidence as possible on the extent to which translation may constitute a propagating factor in language change, rather than attempting to establish a conclusive causal relationship.

Beyond establishing the existence of translation-induced language change, some researchers have examined the factors which may encourage or inhibit such changes from occurring and taking hold. Outlining a tentative typology of translation-induced language change, Kranich, Becher and Höder (2011) put forward a range of key factors which play an important role in the manifestation of changes, with the underlying assumption that translation-induced language change does not differ markedly from other types of language contact. In this way, they highlight 'intensity of contact' and 'length of contact', as well as 'prestige of SL' and 'typological proximity', among other factors, as particularly relevant in the realisation of changes and their adoption over time (pp. 18). In addition, 'cognitive pressure' and the complexity of tasks involved often cause speakers 'to employ word-forms from the 'outside'

language on a regular basis when interacting in the ‘inside’ language.’ (Pang & Wang, 2020, pp. 433). In light of this, this thesis examines a language context which is a propitious site for language change, since the multilingual EU translation environment ensures significant intensity and length of contact between the languages in question, which also have relatively high typological similarity and involve necessarily complex tasks under high levels of cognitive pressure (see Clay, forthcoming). Although they constitute an excellent resource for research of this kind, this will be the first study to use translations of EU legal texts as part of a wider investigation into translation-induced language change.

In order to better define this process in translation-induced language change, Malamatidou (2016) applies some of the key concepts from the field of contact linguistics, namely Johanson’s Code-Copying Framework (1993, 1999, 2002), in light of the observation that translation is not significantly different from other linguistic activities and can therefore be described using the same mechanism. According to this framework, contact-induced changes occur along a continuum, beginning with sporadic ‘momentary copies’, before becoming ‘habitualised copies’, where linguistic features or patterns start to be used frequently by a specific group, in this case translators, terminologists and lawyer-linguists in the EU institutions (Johanson, 2002, pp. 298). When these copies become fully integrated into general usage, they can be considered as having become ‘conventionalised’ (ibid. pp. 299). This framework can also be usefully applied to the results of this thesis in order to gain a clearer understanding of the various stages of development which any detected contact-induced language change may have passed through and which stage it may currently be at.

As demonstrated in this section, previous research has provided a substantial amount of evidence to show the capacity of translation to propagate language change in certain circumstances. It has also offered insights into the factors promoting or inhibiting such changes. This research forms the basis on which this thesis seeks to build in seeking to contribute an additional body of evidence to the current discussion on translation-induced language change and understand its effects in the field of migration in EU law and press articles.

2.2 Legal Translation in the EU Context

Since a key aspect of this research project on translation-induced language change in the field of migration explores legal translation in the EU context, it is important to closely examine the particularities of legal translation, especially in light of the unusual translation situation which exists within the EU institutions. Legal translation broadly refers to 'translation of the law and other communications in the legal setting' (Cao, 2007, pp. 12). Put simply, the task of the legal translator is to use a legal text in one language to create an equivalent legal text in another language such that a legal decision-maker will arrive at the same conclusion irrespective of the language version used (Ainsworth, 2014). Within translation studies, translation of the law is considered its own discrete category due to the unique challenges associated with it, which are not shared by other areas of translation for specific purposes (Felici, 2010). It goes without saying that the practice of legal translation is far from straight-forward. Firstly, legal language in general is characterised by a number of features which make translation of it especially difficult. These include extremely long sentences (Tiersma, 1999), highly specialised terminology to convey specific

knowledge (Rouski, 2013), terms belonging to general language but whose meaning has been expanded (Cornu, 1990), complex syntax and structures (Cao, 2007), highly formulaic language (Mattila, 2006), opaque language and highly particularized forms of discourse which are difficult to interpret (Ainsworth, 2014), and a mixture of very archaic features and innovative usage (Tiersma, 1999). The legal translator therefore needs a clear understanding of complex legal concepts and a sound grasp of how and why legal professionals write in the way they do, so that they can adopt the same sensitivities of language when translating legal texts. The issue of specialised terminology and how this terminology is translated across different languages is especially interesting in law, where terms are created 'not through usage, but by authoritative pronouncements of courts or legislatures' (Tiersma, 1999, pp. 109), and then directly applied to concrete scenarios, making it a particularly interesting area of study in this research project.

However, as well as these purely linguistic challenges, legal translation poses an even greater challenge as a result of the way in which language and the law are so inextricably intertwined. Indeed, since law exists and is expressed *through* language, 'law cannot be understood without an appreciation of the nature of language' (Ainsworth, 2014, pp. 53). However, since legal systems and cultures have developed in different ways, there is never perfect equivalence or overlap between them. The translator's already difficult task of achieving a level of equivalence between the source and target texts, where 'the semantic domain of a word in one language very rarely overlaps exactly with the semantic domain of its closest equivalent in another language' (Doczekalska, 2009, pp. 338) becomes yet more daunting with the introduction of incompatible legal systems. Šarčević identifies the 'incongruency of legal systems' (1997, pp. 13) as the principal challenge to the legal

translator, who is tasked with producing a legal text in the target language which accurately reflects the content of the source text while also respecting the linguistic norms of the target language *and* the legal conventions of the target language legal system. This is especially apparent when translating between common law and civil law systems with entirely different legal histories and cultures. Legal translation is therefore characterised by ‘a strong conflict between accuracy and naturalness’ (Biel, 2014, pp. 43). This is not purely of theoretical importance, since legal texts ‘have an inescapably instrumental character: they exist to actually do things that have concrete, dramatic, often far-reaching consequences in the real world’ (Ainsworth, 2014, pp. 47). The legal translator is therefore obliged to seek ways to map the source text meaning onto a target language with a different legal reality and culture.

With regard to terminology, due to the ‘system bound nature of legal terms, full equivalence between terms from two legal systems is rare’ (Biel, 2014, pp. 42), forcing the translator to deal with only partial equivalence or ‘conceptual voids’. Legal language differs from other specialised fields, such as medicine or mathematics, in that each of the world’s legal systems has its own unique legal terminology and concepts, which differ from legal systems in other countries (McAuliffe, 2016). For example, the concept of legal adulthood varies dramatically between different countries, meaning that in some jurisdictions, the term ‘minor’ is applied to a person until the age of 21, while in others they enter adulthood by the age of 16. The methods and techniques used by translators to overcome this problem of system-bound legal terminology play a central role in this research project, since contact-induced language change often occurs where languages have terminological gaps or new concepts emerge and require labels. Leung (2014, pp. 58) describes the idea of

equivalence in legal translation as ‘legal fiction’ which can be understood as ‘purposeful make-believe, consisting of explicit or implied statements that are not literally true’. When discussing treaties in two or more languages, the Vienna Convention on the Law of Treaties (1969) states that ‘the terms of the treaty are presumed to have the same meaning in each authentic text’, although whether such perfect legal uniformity can be achieved in practice remains a contentious issue. As discussed below, this point becomes especially relevant for EU legal texts, where each language version is deemed to be equally authentic. In this regard, Legrand argues that ‘law is a cultural phenomenon…therefore, differences across legal cultures can only ever be overcome imperfectly’ (2005, pp. 42).

Likewise, deviations in the makeup of different languages lead to different linguistic preferences ‘based on the patterns available to express what we experience’ (Filipovic, 2013, pp. 340). For instance, ambiguities – a common feature of legal texts – are liable to appear in the translation of legal documents due to differences in features such as word order and phrasal structure (Solan, 2014). Moreover, since legal translators often lack the background in which the document has been created, it is not always evident whether obscurity in a text is inadvertent or a result of carefully negotiated agreements between the parties involved (Correia, 2002, pp. 42).

Consequently, there is a range of reasons why legal translation is particularly fraught with difficulties. The specific legal texts analysed in this research project are English, French and Italian versions of documents produced by the EU institutions, which introduce further complexity into the picture. This complexity stems from the exceptional approach the European Union takes to multilingual law-making, which it is important to explore in detail in order to fully contextualise the results of the corpus

analysis carried out here. The European Union is a unique multilingual law-making environment which has undergone huge developments since its founding, originally as the European Coal and Steel Community, back in 1951. Each EU institution has its own decentralised translation service customised to meet the needs of that body, and the largest translation bureau is the Directorate General for Translation (DGT) which serves the European Commission.

The multilingual system in place today within the EU defies comparison with any other national community or confederation of states in its complexity and singularity (Tosi, 2007). This makes legal texts both a fascinating object of study for this research, but also inevitably introduces further levels of complexity into the analysis process. It is the only international organisation to grant official language status to the official language(s) of almost all its Member States[2] and the number of official languages has expanded greatly from the initial four to the 24 which currently enjoy this status. However, despite this significant expansion, the principle of equality in legislation between these languages has remained constant, as first set out in EEC Council Regulation number 1 (1958), 'determining the languages to be used by the European Economic Community', which has been consistently updated with each new accession, with the latest amendment being made in 2013 to establish a current total of 24 'official languages and working languages of the institutions' (Council Regulation (EU) No 517/2013, 2013). This means that all EU citizens can access legislation in one of their country's official languages and can communicate with the EU institutions in any of its official languages. EU translation is therefore an example of institutional translation which is underpinned by the central principle of allowing all citizens, and public and private entities to be able to communicate with EU

[2] with the notable exception of Luxembourgish and Turkish for Cyprus

institutions and have undisputed access EU law, regardless of their native language[3] (Biel, 2014; Šarčević, 2013).

One of the goals of the European Union, through its foundation treaties and its huge array of secondary legislation, is to establish a single system of law, usually referred to as EU law (Robertson, 2010). This secondary legislation consists of binding regulations, directives and decisions, and non-binding legal instruments in the form of recommendations and opinions (Article 110, *Consolidated version of Treaty establishing the European Community*, 2002). In order for this goal to be achieved, as mentioned above, all the language versions of all legal documents produced by the EU are automatically deemed to be equally binding or equally authentic (Cao, 2007). Indeed, the concept of 'authenticity' of all language versions is emphasised repeatedly, insofar as all EU legislative texts become equally valid and authoritative at the moment of their official publication (Felici, 2010). When this legislation is adopted, 'such texts are not mere translations of the law, they are the law itself' (Šarčević, 1997, pp. 20). Indeed, it is due to this concept of equal validity that the EU makes no reference to 'original versions' and 'translations', since this would imply a hierarchy of authenticity, which would violate the underlying principle of equality. Instead, the EU maintains that legal documents are drafted in all languages simultaneously, and that none of the versions are derivative from any other (Koskinen, 2000). A logical consequence of this approach is therefore that all authentic texts in all the EU's official languages should be presumed to have exactly the same meaning (Šarčević, 2013). This strict equivalence between the various language versions means that they carry the same legal value and can be invoked

[3] In addition, Council Conclusion 2005/C 148/01 (2005) permitted the use of additional regional languages within the Council and possibly other Institutions and bodies of the EU.

by EU citizens or businesses regardless of their Member State of origin (Correia, 2002).

However, despite this rigidly held and frequently stressed official EU position of equal authenticity, in practice, the reality of EU legal text production is entirely different. In spite of claims of linguistic equality, it is often conceded that, in fact, ‘behind de jure equality there is also a de facto inequality’ (Fischer, 2009). Given the volume of languages involved, simultaneous, multilingual drafting is simply not feasible, meaning that the majority of documents are drafted first in one language, and are translated into others afterwards (Fischer, 2010). In the past, most of these original drafts were written in French. In 1994, a survey carried out among EU officials showed that 59% were using French as an original drafting language, while 33% were using English (Mattila, 2006, pp. 33). However, the ineluctable rise of English in the institutions meant that by 2014, 81.3% of all legislation was being originally drafted in English (compared to 77.6% in 2012 and 62% in 2004), followed by French, with only 5% (Balič, 2016, pp. 132). A recent report evaluating multilingualism in individual institutions found that 95% of documents produced by the General Secretariat of the Council had English as the source language by 2018, while 85.5% of the documents issued by the Commission had English as the source language (Lequesne, 2020, pp. 46-50). A pecking order of languages has been created by the European Commission by identifying English, French and German as ‘procedural languages’ – the languages in which official documents need to be provided at the initial stage (Baaij, 2018, pp. 61). Equal status for all other languages is provided later through authenticated translations (Felici, 2010). It appears that in the production of EU legal texts, the ideal of linguistic equality has therefore been trumped by the practical reality of producing parallel documents in over 20 different

languages. Indeed, over the course of the last few decades, it has become increasingly clear that ‘behind the cover of equal authenticity of EU languages, the English language version has become the informal and de facto source text from which EU’s multilingual legislation originates’ (Baaij, 2018, pp. 7). This observation is critical for the purposes of this thesis because it recognises that despite the principle of equality, translation does indeed take place on the basis of an original draft version. It is therefore possible, at least in theory, for translation-induced language change to occur in this context.

It is nonetheless important to take into account the various steps in the EU translation process, which is much more complex than a standard translation process. Despite the recognition that the different language versions of the legal texts originate from an initial draft document in one language, the source text-target text relationship familiar to most translation practices is significantly more complicated in this case. The legal drafting starts with a planning stage to work out a set of priorities followed by an impact assessment (Čavoški, 2020). The initial draft of legislation is written up by technical experts, who are not usually drafting experts or lawyers, in the Commission Directorate-General for the subject in question (Robinson, 2014). Since these initial drafts are mainly written in English, the inevitable consequence is that, most of the time, the drafting is carried out by non-native speakers. This inevitably leads to initial drafts with various non-native lexical and syntactic elements, which are then examined and revised by lawyer-linguists at the Commission to ensure that the language is ‘clear, precise and translatable’ (Šarčević, 2013, pp. 9). In parallel, during this drafting process, the Legal Service of the Commission supports internal analysis of scientific evidence and working groups with Member States and subject experts at technical and political levels (Čavoški,

2020). This is usually done on the basis of one language version (usually English) for reasons of convenience and expediency, and the other language versions are subsequently adapted to fit with the base language version (Robertson, 2010).

Only at the end of each legislative stage do translators render the relevant draft proposals, amendments and final instruments into the other official languages (Baaij, 2018). It is not feasible for each and every stage of the drafting process to include extensive translation and interpreting, so a large part of the internal work is conducted in a language mastered by all staff involved, which, for the overwhelming majority, is now English (Robinson, 2014). Indeed, due to the high number of languages which now have official status in the EU, simultaneous drafting has simply become impossible (Otero Fernández, 2020, pp. 51). In many cases, the institutions establish ‘document chains’ where each new document is anticipated by previous documents, and also paves the way for new documents which take the issue to the next level of development (Koskinen, 2008, pp. 133). In this multi-layered legislative drafting process, from the genesis of the legislation, and through each of the drafting stages, it is clear that there is a de facto base version – usually English – which is consulted by the working groups and experts negotiating the formation of the legislation (Clay, forthcoming). This ultimately means, for the purposes of the research presented in this thesis, that the hypothesis of translation-induced language change being investigated can be legitimately applied to EU legal texts. This results from the fact that the process involves a base ‘source text’ which the other versions are derived from through translation, albeit in a much more complex manner compared to the standard translation process.

Nonetheless, it is important for the design of this research study to take account of the somewhat elusive nature of the source text in the EU context. Indeed, the

concept of source text as usually understood in translation studies is problematic in EU translation due to the ‘complex multistage (5-10 stages) drafting process intermingled with translation at all these stages, with constant switching of languages, rewriting and changes to draft documents’ (Biel, 2019, pp. 8). Occasionally, the two ‘pseudo’ source languages may be used within the drafting of the same document (Felici, 2010, pp. 102), with some parts being written in English, and others in French (Tosi, 2007). The process is not entirely one-way, since the source language text is compared with the other language versions and adjusted where necessary to reduce ambiguities, resulting in the language versions becoming ‘fused’ together into a unified message (Robertson, 2010). The language is amended throughout the process to reflect the difficult compromises reached between Member States and the EU institutions. Dollerup (2004, pp. 197) characterises the source text as ‘a fluid and changeable mass of text’ made up of a collection of different elements. It has thus been argued that research in parallel corpora or terminology in this area is difficult ‘since the notion of source language and source text is rather vague’ (Fischer, 2010, pp. 25). However, researchers in this field are increasingly adopting corpus-based approaches to gain new insights into legal translation practices in multilingual jurisdictions (e.g. Prieto-Ramos, 2019; Trklja & McAuliffe, 2019).

While there can be no question that the source text-target text relationship is less straight-forward than in a normal translation process, this research project seeks to demonstrate that corpus analysis methods can be used to examine terminological variation in EU legal texts with a view to exploring the influence of translation on language change. Moreover, previous studies of this kind have been limited to using relatively small corpora only including texts for which experts from EU translation

departments were able to testify as to the genesis of the source and target text in every case (e.g. Kerremans & Temmerman, 2016). Since the approach in this thesis does not make any a priori assumption about translation direction, it is able to use a larger collection of texts, instead using the data derived from the analysis to draw conclusions regarding the identity of the source language(s).

To this end, the methodological design of the project, as set out in chapter 3, starts from a point of equality between the languages studied, which does not assume a source-target relationship, but instead seeks to provide evidence of such a relationship in its results, showing that the influence of the language used in the initial drafts and in the negotiations of working groups can still be detected despite the multi-layered drafting process. After all, it remains the case that 'a single language version, *of one kind or another*, is *always* the starting point for the production of the others by the act of translation' (Correia, 2002, pp. 40) [My italics]. Therefore, despite these atypical aspects of the translation process, it is still possible to maintain that elements of the de facto source language could still influence various linguistic features of the translated versions, which could ultimately result in observable translation-induced language change.

Researchers exploring the language of the EU have argued that the continuous translation of large volumes of legal texts within the EU has led to the emergence of new, hybrid varieties (Mori, 2018a). Referred to as 'eurospeak' (Robertson, 2014, pp. 160) or 'eurolects' (Goffin, 1994), among other labels, these varieties have patterns of usage that differ from standard varieties (Mori, 2018a). These alternative patterns are partly the result of intensive multilingual contact, which is further reinforced by the translation approach taken by the EU institutions. One of the EU's translation guidelines to be followed is the requirement to maintain the same sentence structure

in all language versions insofar as is possible (Čavoški, 2017). This means that the various language versions end up being mirror images of each other, with sentence by sentence and paragraph by paragraph correspondence, with no regard for the textual norms of the languages involved. This inevitably imposes restrictions on the translators, who are obliged to stick rigidly to the source text structure (Sosoni, 2012, pp. 84). Indeed, the quality guidelines issued by the DGT, discussed further below, admit that there is 'limited leeway of localising' to target language conventions (European Commission, 2015, pp. 14). This stringent requirement means that the textual and syntactic structure of many language versions of EU texts must therefore be strongly influenced by the structure of the de facto source language(s) (Clay, forthcoming). For instance, it has been observed that the Slovak language version of EU texts 'imitates the English source text' and that the source text 'represents a language matrix which is filled with the Slovak content' (Bednárová-Gibová, 2017, pp. 154). As a result, this approach provides ample opportunity for structural features of terms used in EU language versions to undergo adaptation through the translation process.

In addition, new terms constantly need to be created in the EU context in order to find terms for EU autonomous concepts, to account for new supranational realities (Doczekalska, 2009) and to avoid borrowing national terminology to convey EU concepts, which may lead to ambiguities (Bajčić, 2011). In most cases, terms are created in the dominant languages (mainly English, French and German) and are subsequently translated into others, meaning that 'target terms (secondary terms) are created on the basis of a source term (primary term), by translation' (Fischer, 2010, pp. 28). The choice of which target term a translator uses for a given concept must therefore be inevitably influenced by the primary concept that has already been

created (Clay, 2022). This is significant for this thesis because it shows how terminology in one language can have a direct impact on terminological choices made in other languages. Such impacts could then manifest themselves in permanent terminological changes through translation if the terminological choices in the target language(s) are consolidated. Moreover, terminology in EU law is still in a state of flux, with some branches of EU Law still lacking their own supranational terminology (Čavoški, 2017). This makes it even more susceptible to variation and change before more stable terminology is established.

The translation units in the EU institutions have various systems to ensure terminological consistency and improve translation quality. The *Interinstitutional Style Guide*, first produced in 1998 and frequently updated, aims to ensure a high degree of standardisation, and contains clear guidelines on standard formats, structure and clear drafting (Publications Office of the European Union, 2022). The interinstitutional terminology database IATE (Inter-Active Terminology for Europe) has also been fully operational within the European Commission since 2005. The objective of this database and various other specialist glossaries produced by the European Union is to improve and maintain terminological consistency within and between different languages. Moreover, the Terminology Coordination Unit at the European Parliament (TermCoord) was specifically created to support translators in carrying out their day-to-day duties by facilitating terminological research and management and increasing the Parliament's contribution to the IATE database (Mariani, 2018). Probably the most important system now used by translators are Computer-Aided Translation tools (CAT tools), such as SDL Trados Studio and a range of other specialised software. These tools are used to store translation memories of previously translated texts, which can then be reused when similar or identical terms and passages are

identified in subsequent texts, thus greatly improving a translator's productivity and consistency. However, another result of these tools is that, once used, terms, phrases, sentences and even whole paragraphs become fixed and standardised. EU text construction then starts to resemble "lego-type" construction through the use of memorised blocks of text (Robertson, 2014, pp. 162). As a result, any changes that may have been introduced to a language version or terms created on the basis of a de facto source language version are quickly propagated, therefore potentially amplifying the effect of source language influence and acting as a catalyst for translation-induced language change in the longer term.

To conclude, translation in the EU context is significantly more complex than the standard translation process. The underlying principle of equal authenticity means that all language versions of the same legal text are deemed to carry equal weight and no single version should be upheld as the source for the other language versions of the same text. Rather than starting a translation from a clearly defined and completed source text, the EU engages in a multi-layered process of intertwined drafting and translation involving all the official languages. In spite of these added layers of complexity, it seems that there is still clear potential for translation-induced language change to occur through EU translation due to the practical realities which lie behind the theoretical principles of equality. Initial drafts of legislation are written now overwhelmingly in English and intermediate negotiations of the texts during the legislative drafting process tend mainly to take place in English too[4]. An unavoidable consequence of this is that 'structure and terminology are set by the base language version and other languages adapt to them as well as they can' (Robertson, 2011).

[4] The language system and translation process at the European Court of Justice differs markedly from the other EU institutions and texts produced by the Court are therefore not included in this analysis. This decision is discussed in more depth in chapter 3.

Furthermore, due to these practical limitations, ‘conceptual thinking’ is also carried out in these dominant languages, increasing their influence on other languages versions yet further (Fischer, 2009, pp. 73). Indeed, it appears that EU translation offers an ideal setting to observe the linguistic effects of translation, by providing as it does a ‘sui generis language contact scenario’ (Mori, 2018a, pp. 9). On a terminological level, terms tend to be established in the dominant languages, and subsequently translated into the others. Moreover, the ‘narrow interpretation of fidelity’ combined with tight deadlines and work pressures also means that translators tend to remain as close as possible to the form of the original, unless the result would be unacceptable (Trim, 2002, pp. 19), which therefore also allows the linguistic features of the original to exert an influence through translation. While the methodology of this study must take into account the fact that the source language of EU legal texts is not clearly discernible, it appears clear that translation in the EU context can in theory act as a site of contact-induced language change.

2.3 The Language of Migration in Law and the Print Media

This thesis specifically examines the terminology used in both legal texts and the press to express a range of concepts in the field of migration. Examining such language is particularly important since the terminology used to describe people and situations can have tangible, real-world effects on individuals’ lived experiences. It is therefore important to reflect on the use of language in this domain in both legal and media contexts and the potential impact different linguistic decisions can have. Terminological choices are extremely important and sensitive when discussing migration as they can ‘imply causality, generate emotional response, and transmit

symbolic meanings' (Bauder, 2014). As indicated, for instance, by the United Nations High Commissioner for Refugees when discussing the use of 'migrant' and 'refugee', 'choices about words do matter' (Edwards, 2016). The wide range of migration-related terms used in both contexts are often anything but neutral. The power of language in this area can have profound effects on how people on the move are perceived, how they see themselves and how they are treated by the law and by host communities. After all, at its core, migration is a phenomenon which deals with the mobility of human beings, so any language describing it and its resulting consequences have a direct impact on its main actors, who, by definition, are usually in a situation of vulnerability (Kostas, 2017). As Goddard and Wierzbicka observe, 'people speak with words, they think with words, they "do things" with words; to a significant extent, words shape people's lives' (2014, pp. 2). Legislatures, policy-makers and media organisations all play a central role in deciding and developing the terminology to be used in relation to the issue of immigration, which is then often propagated among and adopted by general speech communities. As this research project examines terminological use and variation in law and the media, it is important to understand the significance of terminological decisions in both contexts and explore previous research which investigates the use of language around migration, as well as gaining an insight into the interaction between these two contexts.

Understandably, the terminology used in legal texts carries significant weight in shaping language use as a result of the authority with which it is invested. In the specific domain of migration, the labels applied to the people and processes involved can have major implications on the legal and moral obligations receiving countries feel towards them, as well as on the way legislation governs the status of refugees

(Sigona, 2018). Moreover, it has been argued that the use of legal terms by supranational organisations, such as the European Union, in its political and communication settings can strongly influence the language used by European citizens (Mariani, 2018). The language of EU law also carries significant influence beyond its own borders, often being consulted and used as a template in other jurisdictions, such as accession countries (Čavoški, 2018). Indeed, legal texts in this area have the power to shape the wider discourse around migration, influencing how citizens and the media consequently understand and perceive it.

A common observation regarding terminology in legal texts on migration is the tendency to use language which dehumanises the people migrating. It is argued that the widespread use of standardised labels for migrants enables them to be seen as 'merely an apolitical set of bureaucratic categories' (Zetter, 2007, pp. 174), and positions them 'within a detached legal discourse', attributing to them a fixed status which would never be assigned to people deemed full members of society (Baker, 2020, pp. 8). It is doubtful that the existing legal terminology used to categorise people on the move is sufficiently nuanced to capture the 'complex and messy social realities' which are inevitably involved in migration (Crawley & Skleparis, 2018, pp. 50). Legal categories used tend to obstruct the nuanced reality of most migrants' situations and 'implicitly convey judgements linked to nationality, race, profession and class, as well as motives, trustworthiness, rights and the overall 'desirability' of migrants', without any regard for the daily lives or perceptions of the migrants themselves (De Haas, Castles & Miller, 2020, pp. 21).

Previous corpus-based studies of legislative texts from different jurisdictions highlight the detachment devices used when addressing the subject of migrants, such as the passive voice, impersonal syntax and extensive use of jargon, along with

depersonalised, neutral lexical items (Loupaki, 2018), and find that migrants are constructed as largely homogeneous groups of people through extremely restricted vocabulary, with the incremental effect being that these people are presented as ‘partial individuals’ whose main aim is to comply with the administrative regulations in place (Pérez-Paredes, Jiménez & Hernández, 2017, pp. 99). Ieracinto and Vigneri, for instance, argue that the European Agenda on Migration produced in response to the ‘human tragedy’ occurring in the Mediterranean tends towards ‘de-historicizing depersonalisation’ which takes little heed of the accounts of the migrants themselves (2018, pp. 77).

Consequently, it is through such dehumanising linguistic devices in legal discourse that the individual stories of migrants are exchanged for ‘an imposed and marginalised status’ (Zetter, 2007, pp. 187). For example, governments persist in labelling certain migrants as guest workers or temporary workers even if they have been living and working in the country in question for many years (De Haas, Castles & Miller, 2020, pp. 21). Legal texts also make use of metaphorical language which portrays migrants as guests in terms such as ‘host country’ and ‘overstayer’, thus viewing migrants purely through the prism of their economic worth or need for support (Taylor, 2021, pp. 13). Labels such as this used extensively in legal texts develop their own legitimacy and become an accepted and convenient shorthand (Zetter, 2007). The overall effect of such unsuitable labelling is that it constructs a ‘group of controllers against a set of individuals […] whose individual identities are of no importance’ (Pérez-Paredes, Jiménez & Hernández, 2017, pp. 98). Ultimately, the use of impersonal language in legal texts, in particular EU texts which are the subject of this research, is entirely expected and a standard element of legal language in general (Matilla, 2006) but when it is applied to the domain of migration, the effect is

to marginalise the people migrating and conceal their human nature. In this way, apparently critical bureaucratic practices end up producing ‘highly discriminatory labels designed to mediate the interests of the state to control in-migration’ (Zetter, 2007, pp. 184).

As well as marginalising and dehumanising migrants through labels and homogeneous groups, others accuse legal language of criminalising those people involved in the migration process. The ‘increasing regulatory focus on extralegal migration’ has served to bring about an ‘illegalization’ of migration, thus contributing to the air of moral panic which surrounds the issue (Dauvergne, 2008, pp. 2). This intertwining of criminal and immigration law has created a system with distinct categories of people, namely ‘insiders’ and ‘outsiders’ or ‘legal’ and ‘illegal’ (Stumpf, 2006, pp. 380). It has been argued that use of the adjective ‘illegal’ in reference to migrants in legal texts is not only ‘intrinsically harmful’ and ‘discriminatory’, but also inaccurate from a legal perspective, since it violates legal principles and international legal obligations (Guia, 2016, pp. 37).

This criminalisation marks a shift in the perception of these people, creating an ‘us-them’ distinction and identifying them as transgressor first and migrant second, which in turn makes crackdown measures easier to implement (Dauvergne, 2008). National and supranational legal texts frequently refer to migrants, refugees and asylum-seekers as ‘illegal’ despite attempts to standardise terminology to avoid the stigmatisation which this terminology carries (Perkowska, 2016). Indeed, such labelling ‘indicts a person’s entire existence, rather than an action, as criminal’ (Hoops & Braitman, 2019, pp. 151). It has long been argued that only actions can be unlawful, not people themselves. After all, motorists who exceed the speed limit are never referred to as ‘illegal drivers’ (Lakoff & Ferguson, 2006). This terminology also

heightens the sense of threat which migrants are perceived as posing since it readily conflates them with criminality. Recent EU texts intensify this sense of threat through expressions such as ‘the pressure of thousands of migrants’, ‘high volumes of arrivals’ and ‘unprecedented pressure’, stressing the urgent need for security measures to combat the problem (Ieracinto & Vigneri, 2018, pp. 80).

Efforts have been made to improve terminology in legal texts on migration and combat the negative connotations which much of the language frequently carries. Apparently aware of the impact such terminology can have, Article 7 of the Council of Europe’s Parliament Assembly Resolution 1509 (2006) states that ‘Assembly prefers to use the term “irregular migrant” to other terms such as “illegal migrant” or “migrant without papers” because the term is ‘more neutral and does not carry […] the stigmatisation of the term “illegal”’, while the European Parliament’s resolution 2007/2145(INI) (2009) called for EU institutions and Member States to stop using the term ‘illegal immigrants’ due to its very negative connotations, instead preferring ‘irregular/undocumented workers/migrants’. However, even this usage has been called into question, since ‘irregular’ still implies that the migrant as a person is somehow not ‘regular’ (Bauder, 2014, pp. 329), and can lead to confusion since ‘irregular migration’ is also used to describe migration which occurs at irregular times, as opposed to regular seasonal migration (Perkowska, 2016, pp. 193).

Other efforts to avoid grouping migrants into homogeneous categories involve the development of more nuanced concepts and terms to better capture the complexities of migration, such as ‘mixed flows’ and ‘transit migration’ (Crawley & Skleparis, 2017, pp. 51). Attempts from the EU institutions to humanise the issue of migration have led to an increased use of the words ‘persons’ and ‘people’ instead of ‘migrants’ (Mariani, 2018, pp. 130). For example, the introduction of the term ‘vulnerable

person' into EU legal terminology demonstrates the efforts being made to recognise the human nature of migration (ibid, pp. 178). This research will shed light on the extent to which migration terminology in the legal texts in the corpus created for this study corroborates the above observations. It will also examine whether there are any terminological differences between the languages in question and assess whether efforts to harmonise and improve this terminology is evident in the results.

As well as terminology used in legal texts, this research also examines the language used to discuss migration in the print media. As set out below, previous research has sought to examine the type of terminology used in the press and the potential impact of such language usage. Gaining an insight on terminology use in this context is critical because the print media is extremely influential in affecting attitudes through its language and is central to the public understanding of complex issues such as migration and its related political implications, thus serving as a kind of 'cognitive shortcut' for making sense of events (Greussing & Boomgaarden, 2017, pp. 1762). After all, the press is a cornerstone of our democracies to which millions of citizens turn every day to help them understand the world around them (De Vreese, 2005). The level of complexity involved in migration leaves room for print media sources to shape people's understandings of what it involves and what migrants are like, even if media discourse is in stark contrast to statistical and policy conceptions of migration (Blinder & Allen, 2016). Press discourse tends to have a long-term effect on audiences, which is compounded over time through consistent repetition of messages. As a result, any terminology used by the press has a cumulative effect through use and re-use, becoming embedded in the public consciousness. Indeed, far from providing passive descriptions of events, the news media actively constructs them, creating, adopting and repeating terminology which suits its purposes (Van

Dijk, 1989). In this way, news providers act as gatekeepers in society, colouring the audience's perception of events through selective wording and appropriation of information.

As with some legal terminology discussed above, various studies have demonstrated that print media sources often use terminology which unfairly criminalises migrants and portrays them as a threat. By playing on the 'us-them' dichotomy, media sources cultivate a sense of fear and existential threat connected to the arrival of migrants, especially when they are presented as numerous (Nelken, 2018). In this way, terms often used in some news contexts equate migrants to 'invaders, fraudulent opportunists or […] threats to national security' (Mahrouse, 2018, pp. 35). In addition, migrant men are frequently portrayed as a 'sexual threat' against the purity of the nation (Nelken, 2018, pp. 52). The indiscriminate use of the term 'crisis' to describe migration flows is immediately stigmatising for the migrants themselves and adds unnecessary alarm to the discourse (Krzyzanowski, Triandafyllidou & Wodak, 2018). With reference to those crossing the Mediterranean on unseaworthy boats, Sigona points to the dominant discursive construction of them as 'disguised economic migrants' and therefore 'illegal' (2018, pp. 457). In the case of refugees and asylum-seekers, such criminalising terminology obscures the fact that a person is entitled to seek refuge or asylum under international law (Bauder, 2014). By using terms such as 'clandestine', 'illegal' and 'bogus' in relation to these people, the press has consequently shaped the perception that refugees seeking asylum are effectively engaging in illicit activities (Zetter, 2007, pp. 183). This blurring of terminology describing people fleeing their home countries and seeking asylum often creates imprecise or misleading representations and connects them with negative themes. For example, research on the coverage of migration in the press in Europe shows

how people arriving in the UK from Afghanistan were presented in the British press as a threat to the UK taxpayer, while Italian regional newspapers linked Nigerian migrants with the threat of unlawful behaviour (Taylor, 2014).

Equally, the press frequently uses language which dehumanises migrants and depicts them as faceless masses. In a corpus-based study of media discourse in various countries, Fotopoulos and Kaimaklioti (2016) found that the most common collocate of migrants was not a word but a number, specifying how many people were arriving, being deported or being repatriated, immediately grouping together all migrants into large homogeneous units. The press tends to construct all migrants, including refugees and asylum-seekers, as sharing similar characteristics, backgrounds, motivations for migrating and economic status and often frames their plight through the impersonal prism of dealing with ‘the issue’ (Khosravinik, 2009, pp. 494). Journalists tend to write from the perspective of ‘ingroup members’, which leads them to portray ethnic minority groups in terms of ‘them’ rather than part of ‘us’ (Van Dijk, 1989, pp. 204). As well as dehumanising migrants, the press uses terminology which strips them of any sense of personal agency. They are ‘helpless victims’ with no power to shape their own narrative – things are either done to them or done for them (Baker, 2020, pp. 10). This is reinforced through language which focuses on immigration as a process, thus nominalising the activity of individuals into the abstract noun of ‘immigration’ (Charteris-Black, 2006, pp. 568). The fact that newspaper stories are rarely told from the perspective of the migrant but rather through other witnesses’ eyes, such as rescuers, doctors and volunteers, further enhances the sense of migrants as an inert mass, without contextualising the reasons behind the migration itself (Ieracitano & Vigneri, 2018).

Metaphors are another common feature of language used to discuss migration in the press, which serves to further compound the perception of otherness and of migrants constituting an existential threat. Metaphors of water, particularly involving ‘waves’, ‘floods’ or ‘flows’ lead to negative attitudes equating migration with natural disasters, which policy makers should strive to prevent (Blinder & Allen, 2016; Fotopoulos & Kaimaklioti, 2016). Metaphors relating to nations as ‘containers’ undergoing a build-up of pressure are also prevalent, encouraging people to think of migrants as objects not worthy of empathy (Charteris-Black, 2006). Media depictions of people on the move as less than human are also widespread, including images of the ‘swarming mass’ (Gatrell, 2019, pp. 22), large quantities and elemental forces (Greussing & Boomgaarden, 2017), and animalistic images such as that of the ‘honeypot’ (Taylor, 2014, pp. 369). The media’s frequent use of the word ‘jungle’ to describe migrant camps also invokes inhuman and barbaric imagery (Mahrouse, 2018, pp. 37). Indeed, it appears that the use of metaphor in press representations of migration has become ‘an integral part of the discourse’ (Taylor, 2021, pp. 3). This metaphorical depiction of people on the move fits into the broader context of generally negative and stigmatising discourse around migration and an overall tendency to link migrants with negative themes.

Nevertheless, there are some signs that some sections of the media are starting to acknowledge the influence of its language in this area since, in 2015 for instance, the news channel *Al Jazeera* decided to ban the use of the word ‘migrants’ because of its tendency to dehumanise the people it refers to (Malone, 2015) and the French newspaper *Libération* called for greater care when choosing which terms to use for people on the move (Aprile & Diaz, 2017, pp. 29). The British newspaper *The Guardian* urges the press to refer to them as people rather than deriding them as

migrants, arguing that ‘reporting should be humane as well as accurate’ (Marsh, 2015). Baker suggests introducing neologisms to ‘disrupt the terminological landscape…and unpack widespread classifications’ as another approach to improving migration terminology (2020, pp. 11). These examples point to a raised awareness in various countries of the effect terminological choices in the press can have on the wider society.

In sum, the language used both in legal texts and the press is extremely influential in shaping perceptions and constructing viewpoints. It is also important to highlight that the two spheres are not mutually exclusive. For instance, the growing tendency to elide criminal and immigration law reflects the language of criminalisation in the media that has a major effect on attitudes and constructs categorisations of insider/outsider and citizen/non-citizen, which in turn influences immigration policy, the laws that are passed and the ways that they are enforced (Hoops & Braitman, 2019). Indeed, the mass media is influential in setting political agendas and framing debates around migration (Berry, Garcia-Blanco & Moore, 2015; Martínez Lirola, 2021). It is the same general terminological categories that dominate both policy and media debates, which are not only inadequate for describing the complex processes at work but also perpetuate and reinforce rudimentary binary concepts used to unfairly distinguish between different people (Crawley & Skleparis, 2017). It can be argued that the general ‘uncritical adoption’ of legal and policy categories, which are not always particularly meaningful or nuanced, stands in the way of a better understanding of migration issues in all contexts (De Haas, Castles & Miller, 2020, pp. 21). This thesis therefore uses a corpus-based approach to examine to what extent the migration-related terminology in both EU legal texts and French, Italian and British newspapers corroborates previous research in this area before exploring

the changes this terminology has undergone over time and the potential reasons for such changes. The profound impact of language in this area as explored here highlights the importance of gaining a better understanding of how this terminology emerges and evolves in both the legal and media spheres, and how these two domains interact.

2.4 Terminological Variation

The specific linguistic focus of this investigation is the terminology used to refer to key concepts in migration and the extent to which this terminology changes over time. In contrast to general-use words, terms can be described as ‘a lexical unit consisting of one or more than one word which represents a concept inside a domain’ and terminology as ‘the vocabulary of a subject field’ (Bessé, Nkwenti-Azeh & Sager, 1997, pp. 152/154). Put simply, terminology deals with special languages, in comparison to lexicology, which is to do with everyday language. It is this specialised function of terminology which enables it ‘to gain the precision necessary for expressing restricted meaning’ (Kaguera, 2002, pp. 15). The focus on terms in this research is driven by the hypothesis that terms used to describe concepts are the fruit of deliberate human activity, whereby term formation consists of motivated decisions taken by those developing and using terms (Sager, 1997). In this way, terminology can provide an insight into cognitive processes underlying language. By closely examining the use of terminology in the field of migration, this thesis therefore aims to gain a better understanding of how terms are formed, translated and propagated within this domain across different languages. The principle governing this cognitive process of naming is our desire to organise concepts based

on the criteria or characteristics we consider to be salient (Aymerich, Fernández-Silva & Cabré, 2008), thus shedding light on how we see the world and the way we categorise notions. Indeed, Goddard and Wierzbicka argue that ‘word meanings are in fact among the most sensitive and underused detectors for registering what is going on in the world – political, social, and mental’ (2014, pp. 2). As Fowler (1991, pp. 4) succinctly puts it, ‘there are always different ways of saying the same thing, and they are not accidental alternatives. Differences in expression carry ideological distinction (and thus differences in representation)’. This applies particularly strongly to terminology, even more so than to everyday word formation, ‘since it relies heavily on an awareness of pre-existing models and patterns’ (Sager, 1997).

Concepts are given linguistic expression through terms to make the world intelligible and communication meaningful. Pearson describes concepts as ‘mental constructs to which we assign labels’ (1998, pp. 11). In specialist discourse, these ‘labels’ constitute the terms we use to refer to the abstractions in our mind. When examining terminology, we should distinguish between the concept itself in the physical world or our mind and its linguistic manifestation. Berenskoetter defines a concept as ‘an image formed in the human mind that helps us to generate knowledge about the world by organizing, naming and giving meaning to its features’ (2016, pp. 4). Indeed, conceptual historians have argued that a ‘concept is tied to a word, but it is at the same time more than the word’ (Koselleck, 1972 cited in Olsen, 2012). Concepts can contain unstable and multiple meanings and exist outside language. For instance, the concept of nationality, beyond its purely linguistic form, is also expressed in the physical world through flags, passports and border walls (Clay, 2022). A concept cannot therefore be equated to a concrete and narrow linguistic form, but at the same time, the sphere occupied by the concept cannot be entirely

separated from the linguistic sphere (Olsen, 2012). Indeed, without language, it is impossible to properly communicate or perceive a concept (Temmerman, 2000). By adopting a methodology which categorises terms into conceptual clusters, as set out in detail in chapter 3, this thesis therefore recognises and visualises the relationship between the highest level of abstraction and the labels we use to describe them.

Traditional notions of terminology are based on the General Theory of Terminology (GTT) developed and pioneered by Eugen Wüster. This theory is largely centred around the standardisation of terminology within and across languages, mainly applicable to scientific and engineering texts, advocating the use of one term for one concept (e.g. Wüster, 1968). According to GTT, concepts are clear-cut and unambiguously delineated on the basis of their characteristics and each concept should only be designated by a single term (Felber, 1984). Indeed, this notion of univocity is upheld in GTT as a principle to be actively and deliberately pursued (Wüster, 1991). GTT also adopts an onomasiological approach which 'takes its starting point in a concept, and investigates by which different expressions the concept can be designated, or named' (Geeraerts, 2010, pp. 23). In this way, proponents of GTT address the concept first, before examining the linguistic expression of that concept as used in specialist texts. This traditional approach seeks to convince users of specialised language of the benefits of standardisation and establish terminology as a scientific discipline in its own right (Bajčić, 2017).

However, more recent research has questioned the foundations of GTT by applying rigorous corpus-based studies to terminological questions (e.g. Bertaccini, Massari & Castagnoli, 2010; Fernández-Silva & Kerremans, 2011). Temmerman argues that 'Traditional Terminology confuses principles, i.e. objectives to be aimed at, with facts which are the foundation of a science' (2000, pp. 15). While practical for technical

domains such as science and engineering, GTT fails to capture the realities of how terminology operates in other fields where terms are used to describe more abstract and variable concepts, such as law. Indeed, although traditional terminologists may prescribe the universal application of the ‘one term, one concept’ principle, this principle is unable to describe how specialised language works in practice. This rule would imply that polysemy, where a word has more than one meaning, and synonymy, where two or more words have the same or similar meanings, have no place in terminology, and yet it has been widely acknowledged that polysemy and synonymy are in fact common features of language for special purposes (Aymerich, Fernández-Silva & Cabré, 2008). Peruzzo argues that terms share many of the features which are generally attributed to lexical units in general language, and since these lexical units display a high degree of variation through polysemy and synonymy, it is logical to expect the same kind of variation in terminology (2013, pp. 41). The entire focus of terminology studies has thus shifted from the langue level (the underlying language system) to the parole level, examining samples of language as it is actually used (ibid.). This new approach to terminology has shown that rather than a neat and ordered system, more often than not categories are in fact ‘fuzzy and cannot be absolutely classified by logical and ontological means’ (Temmerman, 2000, pp. 16). Indeed, this approach posits that conceptual categories evolve, perspectives differ, terms change their meaning and our understanding of the world develops (ibid.). There is a range of possibilities available when concepts are named, leading to situations where different denominations refer to the same concept on a cognitive level (Aymerich, Fernández-Silva & Cabré, 2008). As a result, this study takes the reality of terminological variation into account by looking beyond

the traditional notion of a univocal relationship between concept and term to a less restrictive framework, which allows for variation and synonymy.

Terminological variation (in contrast to denominative variation, which examines general lexical units as opposed to specialist terms) can thus be defined as the phenomenon of various terms being used to refer to the same concept, either on an intralingual level (synonymy) or across languages (translation equivalents) (Fernández-Silva & Kerremans, 2011). This variation can be both synchronous, with alternative terms for a single concept existing concurrently, and diachronous, with different terms for a concept emerging and disappearing over a period of time. Rather than taking an onomasiological approach as advocated by GTT, this research adopts a semasiological approach, which takes the linguistic forms i.e. the terms themselves, as the starting point of the investigation from which to construct conceptual categories (Temmerman, 2000). The corpus-based methodology used in this research project promotes the application of the semasiological approach, since the methodology is driven by the terms actually used in written texts rather than taking abstract concepts themselves as the starting point of the study.

A number of previous research studies have sought to both provide evidence of interlingual and intralingual terminological variation and examine the various reasons for such phenomena. According to Freixa's typology of causes (2006), as well as dialectal, functional and discursive reasons for variation, such as geographical variation, level of specialisation of the recipient and avoidance of repetition, it is also demonstrated that terminological variation can be motivated by interlinguistic and cognitive causes, which are particularly important for the purposes of this thesis. Contact between languages is a common cause of synonymy in terminology since it can often lead specialists, either intentionally or not, to opt for a linguistic form from

another language, both for reasons of prestige or communicative effectiveness (Freixa, 2006). This finding fits with the overarching hypothesis of this thesis, whereby contact between languages in translation can ultimately help to propagate terminological changes within those languages.

Moreover, on a cognitive level, terminological instability is also common when there are evolutions of knowledge and thought within a domain leading to novel vocabulary and 'imprecise outlines of concepts' (ibid., pp. 64). For instance, in the domain of telecommunications, Bertaccini, Massari and Castagnoli highlight the general terminological 'chaos' caused by the lack of 'clear and fixed terminology' (2010, pp. 11). This is also relevant to the field of migration law in EU texts, which evolved rapidly during the time periods under investigation in this research, as discussed in chapter 3. It is therefore reasonable to anticipate a degree of terminological variation in this domain, even over a short period, in light of its relative newness in the EU legal context (Picton, 2014). Since the meaning of a word is located in a particular context, it evolves with its context over time, and there is no reason to expect EU terminology to be any different in this respect (Temmerman, 2018, pp. 18). Terminological variation therefore occurs due to a range of motivations influencing the application of terms, whether internal (e.g. the writer's perspective or the communicative situation) or external (e.g. after contact with a term for that concept in another language) (Fernández-Silva & Kerremans, 2011).

Studies examining terminological equivalence across different languages demonstrate that when specialised texts are translated, variation within a language (intralingual variation) tends to be reflected in translations (Kerremans & Temmerman, 2016; Fernández-Silva & Kerremans, 2011). Unsurprisingly, multiword terms are especially susceptible to variation due to their cognitive and structural

complexity, which in turn leads to a wide variety of translation solutions for such terms (León-Araúz, Cabezas-García & Reimerink, 2020). Nevertheless, it is not inevitable that variation in a source language will automatically be reflected in a target language or vice versa. Indeed, in her study on short-term terminological evolution in EU texts, Peruzzo identified 'cases where univocal terms are used in one of the two languages under examination, while in the other non-univocal terms are used' (2013, pp. 42). Some interlingual variation is caused by the fact that the established target language equivalent for a given term deviates from a direct or 'literal' translation of that term, meaning that there is not perfect correspondence between equivalence from terminological and translation perspectives (Kerremans & Temmerman, 2016). For example, Kerremans & Temmerman's research highlights various examples of this phenomenon in the environmental domain, such as the translation of the English term 'air pollution' into French as '*qualité de l'air*' [quality of the air], whereby interlingual terminological variation occurs in order to produce a more idiomatic translation (ibid., pp. 58). Indeed, it would be unreasonable to expect full terminological equivalence between terminology in different languages since this would imply an overly literal translation strategy. Deviations from a literal translation of a term are sometimes necessary in favour of a term or translation equivalent that better fits the specific context in question (ibid.). With this in mind, this thesis also sheds light on the extent to which terminological variation in a text in one language is reflected in translated versions of the same text.

This thesis also addresses the potential of specialised legal terms to transcend genre boundaries and become used in general non-expert contexts such as news articles. This process has been described as de-terminologization (Meyer & Mackintosh, 2000), where terms become of general interest beyond the confines of

the experts initially using them. In this process, a term may either retain its terminological sense or, alternatively, undergo substantial semantic changes and a dilution of its original terminological meaning (ibid, pp. 113). Given the media coverage of migration-related stories in recent years and decades, it is not unreasonable to expect certain terms previously reserved for legal texts on migration to be used in press articles on the topic. One objective of this thesis is therefore to test this theory of de-terminologization in this specific context to verify its applicability to legal migration terms becoming de-terminologized for use in press articles for a general readership.

This chapter has provided a background on various key areas of relevance to the study presented in this thesis. This overview of the existing literature places the research set out here in context and underlines where this study fits in relation to previous work carried out in these fields. The following chapter provides a comprehensive description of the methodology adopted for this study.

3. General Methodology

3.1 Corpus-Based Approach

This thesis adopts a corpus-based methodology in order to provide answers to the central research questions set out in the introduction to this thesis and stated in more detail below. For the purposes of this research, a corpus can be defined as ‘a finite body of text, sampled to be maximally representative of a particular variety of language, and which can be stored and manipulated using a computer’ (McEnery & Wilson, 2001, pp. 73). As a research approach, corpus linguistics can be used to support empirical investigations of language variation and use, and is a particularly useful approach since it provides greater generalizability and validity than would otherwise be feasible through other approaches (Biber, 2015). Indeed, the findings produced by such an analysis can be robustly argued and supported by a series of quantitative and qualitative analyses based on actual patterns identified in pre-existing texts, without relying on any *a priori* assumptions or intuition. The approaches applied in this study are referred to as ‘corpus-based’ since they use the analysis of corpora ‘to describe the systematic patterns of variation and use’ for specific linguistic features, in this case terms, as defined in chapter 2 (Biber, 2009, pp. 169). Corpus methods were initially applied to translation studies in order to identify features and patterns specific to translated texts (Baker, 1993). Over time, the application of corpus methods in translation studies broadened to encompass research into variation in translated texts (Baker, 2004) and translation-induced language change (e.g. Amouzadeh & House, 2010; Malamatidou, 2011), which are the two specific areas that are examined in this research study. The results of recent translation studies research using corpus approaches are discussed in chapter 2. The methodology involves analysis of two purpose-built multilingual corpora,

EUMigrLaw Corpus and *MigrationInTheNews Corpus*, created using the online corpus analysis software Sketch Engine (Kilgariff et al. 2014). The ability to build multilingual, sentence-aligned, parts-of-speech tagged corpora and conduct a wide range of linguistic analyses made this software an appropriate choice for this research. These corpora are described in detail in the sections below along with details on the multi-stage process by which each was created. The originality of the methodology adopted in this thesis lies both in the fact that it examines the effects of translation-induced language change on three languages rather than a language pair, and also in the novel applications of corpus data to gain an insight into terminology convergence and divergence at various points throughout the three-stage analysis.

3.2 *EUMigrLaw Corpus*

The *EUMigrLaw Corpus* is a sentence-aligned, multilingual, parallel corpus containing the English, French and Italian versions of EU legal texts on the subject of migration. For the purposes of this research, legal texts are broadly defined as the entirety of the legally binding and non-legally binding instruments at the disposal of the EU to implement policies and programmes, take measures and issue advice to member states (EU monitor, 2010). While prototypical parallel corpora usually have a recognizable source text, the *EUMigrLaw Corpus* is distinct in that there is no clearly identifiable source text for reasons relating to the EU language regime and translation system, as described in chapter 2. Nonetheless, as also argued in the same chapter, it is legitimate to suppose that translation took place involving the texts in question, either as source or target. This corpus is therefore referred to as a

parallel corpus, but there are no *a priori* judgements regarding the source text in each case. For the purposes of conducting a diachronic analysis, the *EUMigrLaw Corpus* is split into two time periods: 1992-1998 and 2015-2018. While these time periods are not of equal length, the duration of these two periods was determined in part to ensure an equal number of words in each subcorpus in the interests of representativeness and balance, as discussed in more depth below. These timeframes were also chosen because they both cover particularly significant periods with regard to the development of migration laws and policies. The importance of these two periods in the development of EU immigration and asylum law is set out in detail in section 3.3 below. Moreover, research on language change shows that a time span of two decades, as in this study, is a sufficient length for language change to occur and be observed (Labov, 1981). Indeed, variation and evolution in terminology in EU texts has been observed over even shorter periods of time (Peruzzo, 2013).

Table 1 below sets out the text types contained in the *EUMigrLaw Corpus*, along with the number of running words (tokens) each one represents in each time period. Determining corpus size by number of running words or tokens is customary practice in corpus linguistics and refers to any string of letters, numbers or punctuation separated by a white space (Brezina, 2018). Although for purposes of clarity, the general term ‘word’ is used throughout this thesis, all references to wordcounts or number of words refers to running words or tokens defined in this way. The discrepancy in the total number of running words between the languages is due to the inherent differences between the languages as well as their different style conventions. For instance, certain punctuation marks, such as colons, are separated by blank spaces in French, but not in English or Italian, thus creating additional

tokens in the overall count. An explanation of each of the text types contained in the corpus is set out in section 3.2.2.

Table 1: Number of tokens in EUMigrLaw Corpus

Text type	Number of running words (tokens)					
	English		French		Italian	
	1992-1998	2015-2018	1992-1998	2015-2018	1992-1998	2015-2018
Regulations	**0**	**38,752**	**0**	**43,024**	**0**	**39,902**
Directives	**0**	**24,382**	**0**	**28,788**	**0**	**25,650**
Council decisions	**5,599**	**36,465**	**6,546**	**42,909**	**6,080**	**37,550**
Treaties	**69,116**	**0**	**73,696**	**0**	**68,089**	**0**
International agreements	**3,663**	**3,073**	**4,159**	**3,600**	**3,891**	**3,292**
Joint actions	**3,182**	**0**	**3,650**	**0**	**3,392**	**0**
Protocols	**1,887**	**0**	**2,075**	**0**	**2,059**	**0**
Total for legally binding texts	**83,447**	**102,672**	**90,126**	**118,321**	**83,511**	**106,394**
Opinions	**86,432**	**34,713**	**106,737**	**44,507**	**90,984**	**39,052**
Recommendations	**1,492**	**0**	**1,761**	**0**	**1,616**	**0**
Reports	**19,518**	**27,273**	**23,084**	**34,398**	**21,645**	**31,366**
Resolutions	**122,519**	**53,417**	**145,360**	**68,976**	**128,963**	**61,330**
Work programmes	**15,492**	**0**	**20,374**	**0**	**17,555**	**0**
Proposals	**6,552**	**125,981**	**7,315**	**147,688**	**6,916**	**135,575**
Communications	**20,188**	**42,950**	**23,525**	**55,225**	**22,460**	**49,155**
Council conclusions	**614**	**0**	**1,122**	**0**	**1,005**	**0**
Total for non-legally binding texts	**272,807**	**284,334**	**329,278**	**350,794**	**291,144**	**316,478**
Totals	**356,254**	**387,006**	**419,403**	**469,115**	**374,655**	**422,872**
Overall total	**2,429,305**					

A total wordcount of approximately 2.5 million for the corpus was chosen based on previous scholarship in corpus linguistics indicating this to be a suitable target for specialist corpora (Nelson, 2010, pp. 56). This corpus size is also dictated by practical considerations, bearing in mind the total number of legal texts actually available in this field and over these dates. Since the *EUMigrLaw Corpus* contains a large proportion of all legal texts produced by the EU in the field of migration over the time periods being studied (rather than simply a small sample), it provides a representative picture of the total population of these texts, meaning that the 'findings based on its contents can be generalised to a larger hypothetical corpus' (Leech, 1991, pp. 27). The detailed process by which the texts were identified, selected and included in the corpus is set out in section 3.2.3. In addition, the text types included were also carefully chosen to ensure balance between the various strata within the corpus (McEnery & Wilson, 1996, pp. 79). Although it is impossible to create a corpus with equal numbers of words from each different legal text type due to the radically different numbers of texts available for each type, efforts were made to ensure an approximately even spread of binding and non-binding text types, as shown by figures in Table 1 above. In addition, it was important to ensure an even chronological spread of texts in the corpus. As described above, the corpus is divided into two time periods (1992-1998 and 2015-2018) and, as far as possible, efforts were made to ensure a similar number of texts in each of the constituent years for both time frames.

3.2.1 The Development of EU Immigration and Asylum Law

As mentioned above, the *EUMigrLaw Corpus* is split into two timeframes that span two particularly significant time periods in the development of EU immigration and asylum law: 1992-1998 and 2015-2018. It is important to examine how immigration and asylum law has developed in the EU over recent decades to contextualise these time periods and demonstrate their significance for this area of law in the EU system. In the early years of the European project, conceived primarily as an economic organisation, migration was not perceived as either important or threatening, and issues of immigration and asylum were mainly deemed to be internal topics not requiring Europe-wide regulation (Karyotis, 2007; Peers, 2011). International norms, such as the United Nations' Refugee Convention (Convention relating to the Status of Refugees, 1951) and the accompanying protocol (Protocol relating to the Status of Refugees, 1967), generally predate EU law on the area of migration and often came to serve as a point of reference when the EU began to address these areas (Azoulai & De Vries, 2014). As the EU continued to work towards creating a common market by abolishing border controls, the first steps in European cooperation in the fields of asylum and migration were triggered as a result (ibid.). Indeed, it is often considered an example of 'functional spillover' from the common market project, the first major result of which was the Schengen Agreement of 1985, signed by Germany, France, Belgium, the Netherlands and Luxembourg, later supplemented by the Schengen Convention in 1990, which committed these states to the gradual abolition of controls at internal borders (Boswell, 2014). However, it was the Maastricht Treaty (*Treaty on European Union*, 1992), which effectively created a formal intergovernmental system for Justice and Home

Affairs cooperation (Peers, 2011), as part of the 'three-pillar system', building on earlier informal initiatives in the area. With the Maastricht Treaty, the area of justice and home affairs, encompassing immigration and asylum policy, was placed in the 'third pillar', designated for areas subject to intergovernmental rather than supranational competence. Figure 1 provides an illustration of how policy areas were divided across the three pillars following the Maastricht Treaty, with justice and home affairs occupying the pillar dedicated to intergovernmental cooperation.

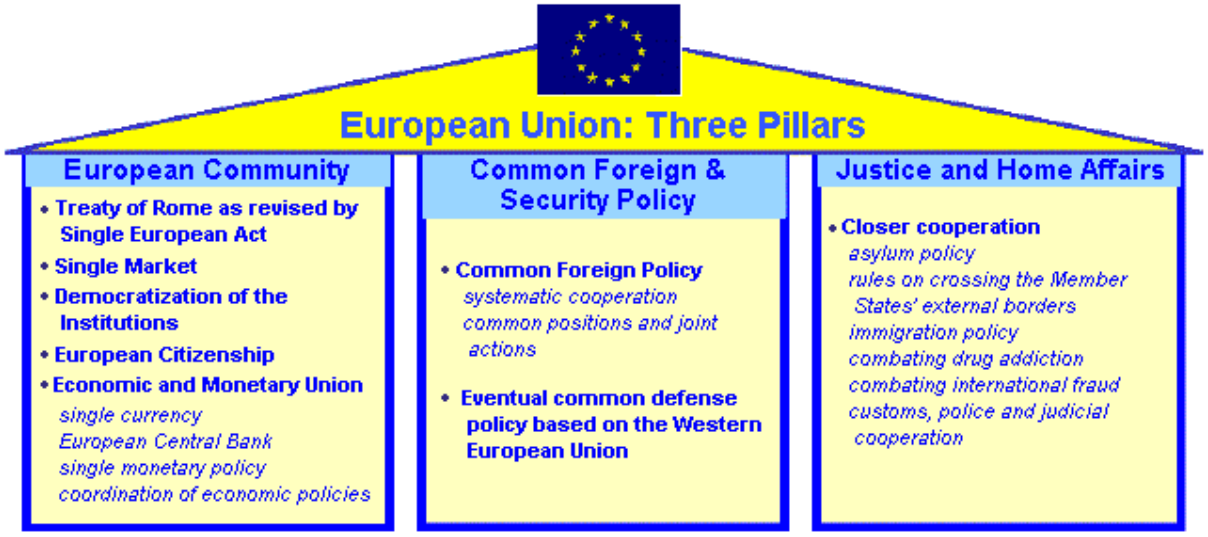


Figure 1: The Pillar Structure of the EU after the Maastricht Treaty (ILSP and MCRC, 2000)

Article K.1 of Title VI of the Treaty, officially entitled the Treaty on European Union, established that 'for the purposes of achieving the objectives of the Union, in particular the free movement of persons', Member States shall henceforth regard the areas of asylum policy, external border controls and immigration policy as 'matters of common interest' (*Treaty on European Union*, 1992). The treaty stipulates that Member States shall 'inform and consult one another' in order to coordinate their action on these matters and shall adopt measures to achieve their common

objectives (ibid.). Due to reluctance and disagreements among Member States over the extent of their involvement, only soft law instruments and policy tools were adopted at this stage (Azoulai & De Vries, 2014; Barnard & Peers, 2020). Nonetheless, the entry into force of the Schengen Convention and the Maastricht Treaty led to a significant volume of new terms entering the EU's migration-related lexicon (Mariani, 2021).

Following on from the Maastricht Treaty, the Treaty of Amsterdam, signed in 1997, amending the Treaty on European Union, significantly increased the EU's powers and scope of activities in the field of migration. It granted the EU competence over visas, asylum, immigration and other policies related to free movement and also contained a 'protocol integrating the Schengen *acquis* into the framework of the European Union' (*Treaty of Amsterdam amending the Treaty on European Union,* 1997) referring jointly to the Schengen Agreement and Convention. This treaty also ushered in the creation of the Area of Freedom, Security and Justice with a view to harmonizing substantive and procedural aspects of refugee law, protecting third-country nationals and coordinating responses to irregular migration (Dauvergne, 2008). Although the continued reticence from Member States and the overly constraining decision-making process meant that many of the agreements reached were simply 'minimal standards' riddled with compromises (Toner, 2014), it was a clear sign that the EU was beginning to develop a common immigration and integration policy covering conditions of entry and residence, along with more extensive provisions on controlling external borders and readmission (Carrera, 2014). The period encompassing the Maastricht and Amsterdam Treaties therefore marks an extremely significant period in the development of migration law, moving from no formal intergovernmental cooperation to an extensive set of provisions on

the matter, thus justifying the choice of this period as the first timeframe for this study.

The period between the timeframes covered by the corpus also saw further development of EU law in the field of migration. The Tampere Conclusions (European Council, 1999) negotiated on the back of the Treaty of Amsterdam aimed to bring together previous piecemeal approaches to immigration and were intended to guide immigration policy and improve integration, as well as show a commitment to a Common European Asylum System (Carrera, 2014; Toner, 2014). The terrorist attacks of 11 September 2001 triggered some highly controversial agreements among Member States around immigration (Brouwer, 2003), effectively ending the liberalization of European migration policy, instead framing the issue as a security threat (Karyotis, 2007). A few years later, in 2004, the European Border and Coast Guard Agency, known as Frontex, was established to assist Member States control their external borders (Barnard & Peers, 2014). The common EU Asylum and Migration Policy implemented in 2005 met the requirement of exchanging information across a broad range of migration-related issues (Mariani, 2021).

Subsequently, large-scale treaty change with the entry into force of the Treaty of Lisbon (2007) further expanded EU competence on immigration and asylum, expanding into the domains of long-term migration, integration and combating irregular migration (Azoulai & De Vries, 2014). The Lisbon Treaty also gave the Charter of Fundamental Rights of the European Union equal legal status as the main treaties, making it legally binding – an especially significant change for rights-sensitive areas such as immigration and asylum (Toner, 2014). Other legislative changes in the area of asylum and migration between the two time periods include Council Directive 86/2003/EC (2003) recognising the right to family reunification,

Council Directive 2009/50/EC (2009) establishing an EU-wide work permit, known as a Blue Card, and Directive 2014/2014/EU (2014) of the European Parliament and of the Council on the rights of seasonal workers. There was clearly significant development in the body of EU migration law in the period between the two subcorpora, and it is therefore reasonable to expect changes to have occurred in the issues addressed as well as in the terminology used to address them. Indeed, it has been observed elsewhere that the number of new terms in the field of migration that entered EU legislative texts over this period, in particular 2000 to 2012, was especially high (Mariani, 2021).

The later time period (2015-2018) was chosen since it covers the period characterised by high levels of inwards migration into EU Member States with hundreds of thousands of people either crossing the Mediterranean or travelling overland through South-East Europe[5]. It was therefore a period during which the EU published a large volume of legally binding and non-binding documents in an attempt to manage the migration situation. This was also occurring against a backdrop of growing opposition in Europe to immigration as part of a wider rejection of a post-national, globalised world, making it more difficult to find a consensus on how to respond to the escalating situation (Abdou, 2020). At this point, Member States were already increasingly facing rebukes from European institutions due to their inability or unwillingness to implement European legislation on migration (Triandafyllidou, 2014).

By late 2013, governments were starting to become aware of increasing migration flows and significant loss of life along the Mediterranean route leading to an

[5] This period has been commonly referred to as the "European migration crisis" or the "refugee crisis", but as has been argued elsewhere, such a description is both 'stigmatizing' and 'unnecessarily alarmistic' (Krzyzanowski, Triandafyllidou and Wodak, 2018, pp. 3), and as such, this terminology is avoided here and throughout.

emergency European Summit being held in April 2015 to work out the fundamental aspects of a common response to the situation (Collett & Le Coz, 2018). The first significant outcome was the European Agenda on Migration, published in May 2015, which set out immediate actions 'in response to the human tragedy in the whole of the Mediterranean', including the establishment of 'hotspots' to help Member States identify, register and fingerprint incoming migrants (*Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, 2015). This was soon followed by a series of new measures, sometimes termed 'externalisation policies', intended to manage the situation, such as Council Decision (EU) 2015/1523 (2015) on the relocation of 40,000 migrants from Italy and Greece and the EU-Turkey Joint Action Plan to support Syrians under temporary protection and strengthen measures against irregular migration (*Report from the Commission to the European Parliament and the Council*, 2016). The largely unexpected arrival of thousands of migrants and refugees into Europe in 2015 and 2016 presented the EU with an urgent predicament in a policy area not designed to manage fast-paced change, which required the rapid deployment of innovative measures and new approaches (Collett & Le Coz, 2018). It therefore marks a key moment in the development of migration and asylum law in Europe and the high volume of legal documents produced during this period in response to the unfolding situation offer a good point of comparison with the earlier timeframe.

3.2.2 Legal Acts in the European Union

The EU corpus contains a representative sample of the legal texts produced by the European Union institutions in all three languages during the two time periods under examination (1992-1998 & 2015-2018). As shown in Table 1 above, it is made up of different types of both legally binding and non-legally binding legal documents which are all freely available for consultation online via the EUR-Lex[6] service in a range of formats and languages. The different document types contained in the *EUMigrLaw Corpus* are summarised in table 2 below.

Table 2: Legally binding and non-legally binding documents included in the EUMigrLaw Corpus

Legally binding	**Non-legally binding**
Treaties	Opinions
Protocols	Recommendations
Decisions	Reports
Regulations	Resolutions
Directives	Work programmes
Joint actions	Proposals
International agreements	Communications
	Council conclusions

Firstly, the corpus contains the text of **treaties** that fall during the time periods. Known in the EU as primary law, they are agreed and ratified by the Member States and set out the objectives and rules of the EU institutions (European e-Justice, 2019). They also set out measures and objectives in main policy areas in which the EU has competence to act either solely, together with member states or where the

[6] https://eur-lex.europa.eu/homepage.html

EU has supportive competences. The corpus also includes **protocols**, which are annexed to treaties and stipulate detailed measures or actions relating to a specific part of the treaty (EU Monitor, 2010).

This primary law and the objectives and principles enshrined in it lay the foundations for the body of law created within the EU structures, which is referred to as secondary law. This consists of binding legal acts: **regulations**, **directives** and **decisions**, all of which are featured in the *EUMigrLaw Corpus*. Article 288 of the Treaty on the Functioning of the European Union (Consolidated Version, 2010) sets out the function of each of these legal acts:

> A regulation shall have general application. It shall be binding in its entirety and directly applicable in all Member States.
>
> A directive shall be binding, as to the result to be achieved, upon each Member State to which it is addressed, but shall leave to the national authorities the choice of form and methods.
>
> A decision shall be binding in its entirety. A decision which specifies those to whom it is addressed shall be binding only on them.

International agreements are another legally binding text type included in the corpus. These are agreements on partnership, cooperation and trade between the European Union and third countries or international organisations.

Recommendations and **opinions** can be categorised as ‘soft law’ insofar as they are not legally binding but can produce some legal effects through more informal ‘soft’ influence. A number of other document types which are included in the corpus also fall into the category of soft law, operating on a hierarchical level below the binding legal acts, including **reports**, **communications**,

resolutions, **work programmes**, **proposals** and **council conclusions**. Although not formally included in Article 288 TFEU above, these instruments have become a significant source of law and give information and guidance on specific areas, often providing updates on the progress of EU actions and the implementation of policies. Inclusion of these soft law documents in the corpus provides a greater depth of insight into the linguistic and policy development of the EU in the area of migration.

In addition, the entry into force of the Maastricht Treaty introduced a number of novel legal acts in this area, in particular **Joint Actions**, which are also included in the corpus. Article K.3.2(b) of the Treaty on European Union states that 'joint actions shall commit the Member States in the positions they adopt and in the conduct of their activity' (*Treaty on European Union*, 1992). However, there was controversy over whether Joint Actions were in fact automatically legally binding in practice, with different Member States adopting differing views on the matter (Peers, 2011). Subsequent reforms ushered in by the Treaty of Amsterdam led to Justice and Home Affairs being moved from the third to the first pillar (see Figure 1). The amendments introduced by the Treaty of Amsterdam meant that these types of acts were gradually phased out and partially replaced with new types of legal acts (Peers, 2011), such as Framework Decisions (Article K.6(b), Treaty of Amsterdam amending the Treaty on European Union, 1997). The duality of legal acts was finally abolished with the entry into force of the Lisbon Treaty, and therefore these particular acts do not appear in the later subcorpus covering the period 2015-2018.

It is also important to note the legal documents which are not included in the corpus. The corpus does not contain any case law produced by the European Court of

Justice. Although the Court's Judgments and the Advocates General's Opinions are available in all official EU languages, the language regime and translation processes differ markedly from the other EU institutions. Indeed, unlike the other EU institutions where all languages are considered to be on an equal footing and no documents are recognised as being either a source or target text, as described in detail in chapter 2, the ECJ has a single internal working language: French. The collegiate judgments produced by the Court are always drafted in French from which the other language versions are then derived (McAuliffe, 2011). Moreover, translations at the ECJ are carried out by lawyer-linguists, whose role differs markedly from standard translators. They are required to combine the roles of translator and legal expert, often bringing their legal expertise to bear on the production of translations in such a way that goes above and beyond usual expectations for a translator (McAuliffe, 2016). The result is that the texts produced by the ECJ are highly distinctive on a textual and lexical level due to the entirely different language regime and translation processes in place (Clay & McAuliffe, 2021), and the legal effects of case law are also entirely different to the other legal texts included in the corpus. Consequently, for the purposes of this research study, they are not included in the corpus. A separate study exploring the use of migration terminology in EU case law could also be conducted using a similar methodology to the one set out here, but this falls outside the scope of this thesis.

3.2.3 Building the *EUMigrLaw Corpus*

Having examined the reasons for using corpus methods, the content of the *EUMigrLaw Corpus* and the rationale behind the content selected, this section systematically sets out the process by which the *EUMigrLaw Corpus* was created.

The first step involves identifying the texts to include in the corpus according to various key criteria. All texts were downloaded from the EUR-LEX website, an online database of EU legal texts that enables multiple criteria to be entered in order to search for and filter the database of legal texts. The first essential requirement for all texts to be included in the corpus is for them to be in the domain of migration. Using the 'advanced search' function, it is possible to search by keyword. For the purposes of this analysis, the texts were searched using the keywords '*immigration*' and/or '*migration*' appearing either in the title or the text. In addition, the date of the documents is another search option that can be adjusted. In this way, it is possible to obtain a list of all texts produced by the EU featuring the keywords '*immigration*' and/or '*migration*' during a certain year. The results are presented in order of relevance, meaning that the texts with the most references to '*immigration*' or '*migration*' appear first. This list can then be manually consulted and appropriate texts selected for inclusion in the corpus taking into account the additional criteria described below. This manual selection step is also important to ensure that the texts selected are all available in English, French and Italian versions.

Once the texts for the corpus have been selected, the three language versions of each text need to be sentence-aligned to enable parallel concordance operations to be conducted. This can be done using the 'LF Aligner' software (Farkas, 2010), which automatically downloads the selected text in .tmx format based on the unique CELEX number attributed to each document in the EUR-LEX database. After entering the CELEX number of a document (e.g 32015D2248) and selecting the language versions required (English, French and Italian in this case), the software then aligns the language versions at sentence level.

When the alignment of the selected texts is complete, the documents can then be uploaded onto the Sketch Engine platform. When building a new corpus, Sketch Engine has the option of uploading pre-aligned texts in order to create a multilingual, parallel corpus. Upon uploading, each language in the source file is processed into a separate monolingual corpus, which is aligned with the corresponding corpora in the other languages. Sketch Engine automatically tags all of the texts in each language with an appropriate tagset[7]. This labels each token in a text to identify its grammatical category, facilitating subsequent analysis. Following these steps, the corpus is ready to be analysed. The general overview of the three-stage analysis process is outlined in section 3.4 below, and the specific methodologies used for each stage of this analysis are set out in detail at the beginning of each analysis chapter.

3.3 *MigrationInTheNews Corpus*

The *MigrationInTheNews Corpus* is a multilingual, comparable corpus of English, French and Italian news texts on the subject of migration. A comparable corpus is a collection of texts selected according to specific criteria (texts on a specific topic, over a given period, from a specific source, etc.) but which remain independent (Kenning, 2010, pp. 487). The *MigrationInTheNews Corpus* does not therefore comprise translated texts but independent news articles in English, French and Italian on the topic of migration. In the context of this study, the purpose of this corpus is to provide an insight into the intralingual terminological change in the field

[7] English: TreeTagger PoS tagset with Sketch Engine modifications – French: French TreeTaggerPoS tagset/French Freeling PoS tagset – Italian: Italian TreeTagger PoS tagset using Marco Baroni's parameter file.

of migration in news articles in English, French and Italian. The *MigrationInTheNews Corpus* therefore serves as a point of reference for the terminological dynamics identified in the *EUMigrLaw Corpus*. As a consequence, with regard to the design of the *MigrationInTheNews Corpus*, a number of key parameters were set in order to closely reflect the *EUMigrLaw Corpus*. Naturally, the time frames covered by this corpus are the same (1992-1998 & 2015-2018) and the total number of tokens in the corpus approximately equivalent (2,363,685 total tokens). Details on the design and creation of the *MigrationInTheNews Corpus* are set out in section 3.3.1. Table 3 below shows the distribution of tokens broken down by language and time period.

Table 3: Number of tokens in MigrationInTheNews Corpus

	English	**French**	**Italian**
1992-1998	381,116	395,661	414,238
2015-2018	381,160	403,303	388,207
Totals	762,276	798,964	802,445
Overall total	2,363,685		

As with the *EUMigrLaw Corpus*, issues of representativeness and balance also needed to be addressed during the construction of the *MigrationInTheNews Corpus*. As discussed in greater depth in section 3.3.1, it was critical to capture texts from a range of different newspapers for each of the three languages represented in the corpus. In addition, the number of words for each individual year was kept consistent so that no single year is over- or under-represented in the corpus.

3.3.1 Building the *MigrationInTheNews Corpus*

As with the *EUMigrLaw Corpus*, having established the key parameters of the corpus in terms of total wordcount and balanced spread of time periods and sources, the first step involves identifying suitable articles to be included in the corpus. The news articles were selected for the corpus using the online news archive LexisNexis (RELX, 2023), which offers a range of advanced search features which facilitate corpus creation. Consequently, articles were searched based on keywords and publication date. Specifically, an advanced search was carried out for articles which included the terms '*migration*' or '*immigration*' and were published between 1 January 1992 and 31 December 1992 (and subsequently for each relevant year in turn). This produces a list of articles in order of relevance i.e. those articles at the top of the list contain the most mentions of the keyword search terms. These articles can then be filtered by publication name, enabling only those newspapers relevant to the purpose of this research to be selected. For instance, English-language and French-language newspapers based outside Europe were filtered out since the purpose of this corpus is to serve as a point of comparison with the *EUMigrLaw Corpus*, containing European-specific terminology. This approach was then repeated for each subsequent year covered by this research.

Following this process, each article in the list is consulted in turn to determine its suitability for inclusion in the corpus, taking into account the criteria set out in section 3.3. In this way, a representative selection of texts can be established for each year and for each language. When this process has been completed, the complete selection of texts can then be uploaded into Sketch Engine. The sets of texts can be grouped by language and uploaded separately to create three discrete subcorpora, together constituting the *MigrationInTheNews Corpus*. As noted above, all the texts

are automatically tagged using the default tagset, making the corpus ready for analysis.

Table 4 below shows the news sources from which articles were selected to be included in the *MigrationInTheNews Corpus* accompanied by a short description of each publication and its political stance[8].

Table 4: News sources in the MigrationInTheNews Corpus by language

English	**French**	**Italian**
• **The Daily Mail** (daily mid-market; right-wing) • **The Evening Standard** (daily freesheet; right-wing) • **The Guardian** (daily broadsheet; left-wing) • **The Times** (daily broadsheet; centre-right) • **The Financial Times** (daily broadsheet; centre-right) • **The Independent** (daily broadsheet; centre-left) • **The Express** (daily tabloid; right-wing) • **The Mirror** (daily tabloid; left-wing) • **The Sun** (daily tabloid; right-wing) • **The Telegraph** (daily broadsheet; right-wing)	• **Le Monde** (national daily; centre) • **Les Echos** (national daily; right-wing) • **La Croix** (national daily; right-wing) • **Le Point** (weekly news magazine; right-wing) • **Agence France Presse** (news agency) • **Le Figaro** (national daily; right-wing) • **L'Humanité** (national daily; left-wing) • **Libération** (national daily; left-wing) • **L'Express** (weekly news magazine; centre-right)	• **La Stampa** (national daily; centre) • **Corriere della Sera** (national daily; centre) • **Il Giorno** (regional daily; centre) • **La Nazione** (regional daily; centre-right) • **La Repubblica** (national daily; centre-left)

[8] Classifications of newspaper types and political leanings for UK publications (Brownlie, 2010; Sirhan, 2021; Blinder and Allen, 2016), for French publications (Lardeau, 2018; Mould, 2011), and for Italian publications (Mancini & Gerli, 2018).

• **The Observer** (weekly broadsheet; left-wing) • **The Scotsman** (daily broadsheet; non-aligned) • **The Metro** (daily freesheet; non-aligned)		

While maintaining the number of sources exactly consistent across the languages would be the ideal scenario, the significant differences between the print media environment in the UK, France and Italy mean that the number of sources available for each varies substantially. It is worth briefly examining the main distinctions between these three media landscapes that explains the discrepancy in the sources available. The UK is characterised by a particularly substantial national press, comparatively much larger than in France or Italy (Firmstone, 2018). The UK press tends to be divided into three sections: i) tabloids, such as *The Sun*, which tend to contain more celebrity and entertainment content; ii) mid-market papers, such as *The Daily Mail* or *The Express*, with a mix of entertainment news and traditional reporting; and iii) broadsheets, such as *The Guardian* and *The Financial Times*, which feature high levels of traditional reporting and news content (Blinder & Allen, 2016, pp. 9). These categories have been used in the descriptions of the UK newspapers in table 4 above. It was important, in order to achieve a representative sample of UK print media, to ensure that newspapers across the three categories were included. In addition, a number of free daily newspapers, known as freesheets, are published in the UK and distributed in large urban areas. Due to their relatively high circulation, these are also included in the corpus. The political stance of UK newspapers is also clearly manifest in their news reporting with highly partisan views being expressed in most publications (Firmstone, 2018). With this in mind, it is also

important to include a broad range of political leanings to ensure that the sweep of terminology that is used in this domain is captured in its entirety so that the corpus is sufficiently representative for the purposes of this research.

The French print media landscape tends to be segmented differently, consisting of daily national and regional newspapers and a periodical press made up of news-based weekly magazines (Lardeau, 2018). The daily 'news and opinion' press in France has dwindled and largely disappeared in recent decades, which may partly explain the fewer publications available in the French newspaper archive, and the remaining dailies have adopted a more neutral tone than that characteristic of UK newspapers (ibid.). Nonetheless, there is still an evident spectrum of political stances represented by the daily newspapers and news-based weekly magazines, which is captured by the various sources included in the corpus, ranging from right-wing publications such as *Le Figaro* and *Le Point* to left-wing outlets such as *Libération* and *L'Humanité*. In addition, the French part of the corpus includes articles from the French news agency, *Agence France Presse,* which strives to provide fact-based and impartial articles on current affairs.

The print media environment in Italy is once again markedly different. Indeed, mass circulation has never been a feature of print media in Italy, and a tabloid press, so commonplace in the UK, is entirely lacking in Italy, with newspapers appealing instead to a relatively elite readership (Mancini & Gerli, 2018). Previous research has pointed to Italy as having a low level of 'newspaper-centrism' compared to countries in northern Europe, with Italians instead tending to consume more television news (Shehata & Strömbäck, 2011, pp. 118), thus leaving less space in the media landscape for the press. This specificity of the Italian media landscape helps to explain the lower availability of sources for press articles to include in the corpus.

Two regional newspapers, *La Nazione* and *Il Giorno*, were also included in the Italian part of the corpus since they have circulations which are not dissimilar to those of the national daily newspapers (Mancini & Gerli, 2018).

The result of these differences in the print media landscapes in the UK, France and Italy is inevitably reflected in the number of sources available for each language on the LexisNexis online news archive used for the purposes of this study. As Nelson puts it, ‘any attempt at corpus creation is therefore a compromise between the hoped for and the achievable’ (2010, pp. 60). Nonetheless, despite this difference, the range of newspapers used as sources in each language is sufficient to ensure that a spectrum of political stances is captured in each subcorpus. Indeed, for each language, there is a mix of news sources which can be positioned towards the left, in the centre and to the right of the political spectrum.

Moreover, in order for this corpus to provide a useful point of comparison against the *EUMigrLaw Corpus*, it was decided that news articles solely referring to migration outside Europe would be excluded from the corpus. This was largely due to the fact that there are large numbers of articles, especially in the later time period, focusing on migration in USA, which tend to use an entirely different set of terms to those used in the European context. While interesting in its own right, this disparity is not directly relevant and falls outside the scope of this research.

The corpus which results from this process naturally covers a very wide range of stories and issues. Many of these include short-term stories of local and regional interest as well as articles discussing more long-term issues of Europe-wide relevance. In terms of specific content, in the earlier time frame (1992-1998), all three languages have numerous articles on the signing of the Maastricht treaty, the

drafting and ratification process of the Treaty of Amsterdam and its implications for migration and the creation of the Schengen Area. The English subcorpora contains a high number of articles on the UK's Asylum and Immigration Act introduced in 1996 and debates around migration prior to the 1997 General Election. The most prevalent articles in the French part are on the anti-immigration laws, known as the *Lois Pasqua*, introduced during this period, as well as stories on the candidates' policies around migration in the 1995 Presidential Election. The most frequently occurring articles in the Italian subcorpora for the earlier time period include those on the entry of migrants and refugees from the former Yugoslavia and the Prodi government's 1997 naval blockade against migratory flows from Albania.

In the later time period, all three languages contain numerous articles on the reaction of European countries and the EU as a whole to the increased migratory flows from the Middle East and North Africa, controversies around the detention of asylum seekers, and the implications of the UK's withdrawal from the EU on migration. Specific to the English subcorpus are a high number of articles on the migration debate prior to the 2015 General Election and the reaction to the 'Windrush Scandal', which involved the government's wrongful detention and deportation of people from the UK. The most prevalent articles in the French subcorpus in the later time period are those related to the 2018 Asylum and Immigration Act and the migration debates during the 2017 Presidential Election campaign. In the Italian part during the later time period, the most common articles refer to the ongoing political debate about managing migratory flows from North Africa and the migrant centres on the Mediterranean island of Lampedusa.

3.4 Step-by-step methodological approach

The overarching methodology of this research project is broken down into three stages which each aim to answer a specific question in order to drive the research to the next stage and inform the next question to be asked. This step-by-step process is set out below, along with the key questions to be answered at each stage. A detailed explanation of each stage is then set out at the beginning of the three subsequent results and analysis chapters (chapters 4, 5 and 6).

The central question which the first stage of the analysis seeks to answer is: **Is there observable intralingual and/or interlingual terminological variation in EU legal texts on migration?** This stage therefore aims to identify variation in terminology, as defined in section 2.4, used both within the individual languages studied and between the different languages in the corpus i.e. when term X is used in language A, is term Y always used in language B? This analysis of terminological variation is conducted both synchronically and diachronically in order to identify variation in terms that occurs concurrently and variation that happens over time. If such variation is detected, the following subsequent question can then be posed: **does such variation indicate terminological convergence or divergence among the languages in the corpus?** This stage uses the notions of terminological distance and interlingual variation index, combined with statistical significance testing to obtain robust, quantitative data on terminological changes. In addition, in order to contextualise the data and address specific examples in the texts, a complementary qualitative analysis is then conducted. The aim of this analysis is to focus in on the direction of any convergence detected i.e. **has language A grown more similar to language B, or vice versa?** This involves analysing patterns of difference and similarity between specific terms used in different languages.

The second stage aims to answer the question: **Is there observable synchronic and/or diachronic terminological variation within each of the languages in migration-related news articles?** The objective of this stage is therefore to ascertain whether and to what extent migration-related terms are used differently in each language both concurrently and over time. This stage examines the relative frequency of the relevant terms and employs statistical techniques including log likelihood and percentage change metrics to show those terms which have increased or decreased to a significant extent over the time period in each language. This quantitative stage also takes into account the percentage each term represents as a proportion of the total number of terms. As part of a complimentary qualitative stage, the patterns of terminological variation observed across the three languages are analysed in order to draw conclusions on similarities and differences between the languages in this context.

The final stage of the analysis constitutes a comprehensive synchronic and diachronic comparison of key migration terms across the two corpora. The key question here is: **are the terminological changes within each language in the *EUMigrLaw Corpus* reflected in the *MigrationInTheNews Corpus* and vice versa?** The methodology of this stage uses a range of different quantitative metrics to detect terminological convergence or divergence between the two text types contained in the corpora. Subsequently, a qualitative analysis looks at individual terms in turn to identify terminological trends across text types and languages. The ultimate aim of this stage is to provide an answer to the question: **are there observable common patterns of change which could point to an influence of the changes in EU terminology detected in stage 1 on terminology in news articles?** Evidence of such could therefore suggest a traceable lineage of influence

from EU legal translation to non-translated news texts. The results of this final stage may therefore confirm or refute the hypothesis that translation-induced language change can occur for migration terminology at the EU level and subsequently become determinologised, crossing genre boundaries to enter into the non-specialist lexicon of news articles.

This chapter has provided a description and justification of the general methodological approach adopted for this study, as well as offering a comprehensive picture of the content of each corpora. Finally, it has set out the key research questions for each stage of the three-part study. The subsequent three chapters (4, 5 and 6) set out the three stages of the analysis, including a specific methodology for each part, followed by results, analysis and discussion.

4. Stage 1: *EUMigrLaw Corpus* Analysis

The first stage in this three-part research study involves the analysis of the *EUMigrLaw Corpus* in order to respond to the initial research questions of this thesis set out above. This chapter describes the specific methodology used for this stage of the study, followed by the results, analysis and discussion of the findings. As stated previously, this section seeks to understand whether there is observable synchronic and/or diachronic intralingual and/or interlingual terminological variation in EU migration texts. If such variation is identified, then this methodology can be applied to understand whether it indicates terminological convergence or divergence among the languages in the corpus. In addition, a qualitative analysis stage is then conducted to ascertain the direction of any convergence detected. The methodology set out for this part of the research study is a development of the methodology presented for an earlier pilot study examining terminological variation on a more limited scope both in terms of the number of languages and terms studied (Clay, 2022).

4.1 Methodology

Since the corpus contains hundreds of migration-related terms, the terms that have been chosen for analysis are limited to those terms referring to people involved in migration (e.g. *asylum-seeker*), at the exclusion of, for instance, verbs (e.g. *repatriate*) and abstract nouns (e.g. *resettlement*). This decision was made both for practical purposes given the huge number of migration-related terms in the corpus, as well as for the fact that terms referring to people in migration are often the focus of scrutiny in both legal circles and the media (cf. Bauder, 2014; Blinder & Allen,

2016; Baker, Gabrielatos & McEnery, 2013). In order to extract all relevant terms to be investigated from the *EUMigrLaw Corpus*, these terms first need to be identified using the wordlist function in Sketch Engine. Sketch Engine automatically produces a list of all the nouns which appear in the corpus in order of frequency thanks to the tagset applied when the texts are uploaded to the corpus, as described in chapter 3. This noun wordlist displays all nouns which appear in the corpus along with their absolute frequency of occurrence. This wordlist can then be manually analysed and all terms referring to people can be extracted.

At this stage, these are all single-word terms, some of which are clearly related to migration (e.g. refugee) and can therefore be included on the keyword list for analysis, while others are much more general (e.g. person) and require further investigation, as described below. Each of these single-word terms, both migration-specific and general, can then be tested using the 'n-grams' function, which retrieves from the corpus all of the relevant multi-word terms featuring those single-word terms. In this study, the n-gram length is set between 2 and 6 words with no minimum frequency, to ensure that all relevant multi-word terms are identified. After all the multi-word terms have been retrieved in this way, those terms which refer to people but are too general to be included in the migration wordlist can be removed (e.g. *young person*), while those related to migration can be included (e.g. *displaced person*). This process can be conducted with each of the three language subcorpora in turn in order to create keyword lists for English, French and Italian.

These lists of single and multi-word terms can then be grouped into conceptual categories. These conceptual categories contain all those terms belonging to the same semantic sphere. For example, the terms found in the English subcorpus '*illegal immigrants*', '*irregular migrants*' and '*irregular immigrants*' all belong to the

same conceptual category, and are therefore grouped under the label IRREGULAR MIGRANTS. The labels for these conceptual categories are provided by the glossary of the European Migration Network. This Network, made up of experts working to offer information on key migration-related issues, provides an up-to-date glossary of around 500 terms reflecting the most recent European policy on migration and asylum, and is also free to consult for both in-house and external translators (Directorate-General for Migration and Home Affairs, 2020).

This stage of the process is similar to the creation of 'formal onomasiological profiles' described by Speelman, Grondelaers & Geeraerts (2003, pp. 319), involving the clustering of sets of alternative words to designate a single concept. The result of this stage is that each conceptual category has its own lexical profile consisting of the different terms used to express that concept. Indeed, it has previously been demonstrated that different terms have been used in EU texts to designate the same referent (e.g. Peruzzo, 2013), thus having numerous different terms within the same conceptual category should not be considered unusual. Table 5 below shows an example for the English conceptual category UNION CITIZENS in the 1992-1998 subcorpus and the various terms appearing in the corpus which fit into this category. (Full results tables for this stage are provided in appendices 1.1-1.6).

Table 5: Terms in the conceptual category UNION CITIZEN for English (1992-1998)

UNION CITIZEN	Citizen of the Union
	EU citizen
	Union citizen
	European citizen

	National of member state

The data gathered by this stage of the methodology enables an analysis of intralingual variation to be carried out. Using the table of conceptual categories and terms within them, the mean, median and mode number of terms per concept can be calculated, where mean represents the average number in a dataset, median is the middle number in a dataset and mode is the most common number in a dataset. In combination, these values provide an indication of the level and nature of the terminological variation present in each language and for both time periods. This information can also be compared across languages and time periods to understand the similarities and differences between the variation identified. These intralingual variation findings can be analysed before moving onto the interlingual variation analysis below.

The next stage involves examining how the respective terms are rendered in the other languages and the extent of interlingual variation present across the languages being studied. Rather than examining all three languages together, this part is broken down into three subsections: English>French, French>Italian and Italian>English. This provides more fine-grained data on the relationships between the various language pairs and will shed light on the terminological distance between the individual languages. In light of the discussion regarding the translation process in the EU, as described in section 2.2, it also means that no one language is deemed as being the source for the other language versions as they are dealt with each in turn. Dealing first with the English term list, the corpus is manually analysed using the parallel concordance function to investigate how each of the terms in English is expressed in the French subcorpus. It is important to reiterate here that, although it

is unavoidable to work from one language to another in this way, no assumption is made regarding translation direction and the same results would be obtained by analysing a parallel concordance from French to English i.e. in the other direction. This is further mitigated by the circular data collection process described above. This data collection produces a table of results showing how each of the English lexical alternatives in each conceptual category is expressed in the French subcorpus, and the frequency with which a pair of terms are used as equivalents. The same process can then be carried out for the French>Italian language pair and the Italian>English language pair. An example of one conceptual category in the English>French language pair in the 2015-2018 subcorpus is shown in table 6 below.

Table 6: Conceptual category for MIGRANT WORKER (English, 2015-2018)

	ENG	FR	Frequency
MIGRANT WORKER	*Foreign worker*	*Travailleur étranger* [Foreign worker]	3
	Foreign worker	*Étranger* [Foreigner]	1
	Foreign worker	*Travailleur* [Worker]	1
	Migrant worker	*Travailleur migrant* [Migrant worker]	19
	Migrant worker	*Travailleur clandestin* [Clandestine worker]	1
	Economic migrant	*Migrant économique* [Economic migrant]	7
	Economic migrant	*Immigrant économique* [Economic immigrant]	1

The next stage in this part of the process is to determine the terminological distance between each of the pairs of terms used to express the various concepts (cf.

Fernández-Silva & Kerremans, 2011). Terminological distance is a systematic way of quantifying the interlingual variation between terms for the same concept in different languages. To determine terminological distance, a value must be manually assigned by the researcher to each pair of terms based on the differences in semantic, morphosyntactic and connotative nature of the two terms. The procedure by which these values are allocated to the term pairs is described below.

A value of '0' is allocated where there is no discernible difference between the terms in the term pair on either a semantic, morphosyntactic and connotative level. Examples of term pairs to which the value '0' is allocated are shown in table 7 below.

Table 7: Examples of term pairs allocated a terminological distance value of 0

Term pair		Terminological distance
(EN) *Migrant worker*	(FR) *Travailleur migrant* [Migrant worker]	0
(FR) *Demandeur d'asile* [Asylum seeker]	(IT) *Richiedente asilo* [Asylum seeker]	0
(IT) *Vittima della tratta* [Victim of trafficking]	(EN) *Victim of trafficking*	0

A value of '0.5' is allocated if there is partial overlap between the terms, but there is some level of terminological divergence. This may be in the form of semantic, morphosyntactic or connotative difference. Examples of term pairs to which the value '0.5' is allocated are shown in table 8 below.

Table 8: Examples of term pairs allocated a terminological distance value of 0.5

Term pair		Terminological distance
(EN) Irregular migrant	(FR) *Migrant en situation irrégulière* [Migrant in an irregular situation]	0.5
(IT) *Straniero* [Foreigner]	(EN) Non-national	0.5
(IT) *Profugo* [See below]	(EN) Refugee	0.5

The first example term pair in table 8 shows morphosyntatic difference, where the syntactic structure of the English term is [adjective + noun], while the French term structure is [noun + prepositional phrase]. The second example presents connotative difference in the conceptual category of THIRD-COUNTRY NATIONAL, since the English term '*non-national*' has neutral connotations in comparison with the Italian term '*straniero*' [foreigner], which often carries negative connotations. Finally, the third term pair in table 8 provides an example of semantic difference. While it is frequently translated into English as '*refugee*', the Italian term '*profugo*' applies to someone who has been forced to leave their country of origin but for whatever reason is not able to request international protection (Greco, 2022), and in strictly legal terms is therefore not synonymous with the term '*rifugiato*' [refugee], as set out in the Geneva Convention (*Convention relating to the status of refugees*, 1951).

A value of '1' is allocated to a pair where there is significant terminological distance between the two terms. In this instance, there is minimal semantic, morphosyntactic and connotative overlap between the terms. Table 9 below shows examples of term pairs to which the value '1' is allocated.

Table 9: Examples of term pairs allocated a terminological distance value of 1

Term pair		Terminological distance
(IT) *Persona in una situazione irregolare* [Person in an irregular situation]	(EN) Illegal immigrant	1
(IT) *Cittadino di un paese terzo* [Citizen of a third country]	(EN) Immigrant	1
(EN) Displaced person	(FR) *Déraciné* [Uprooted]	1

All of the examples presented here show term pairs which have semantic, morphosyntactic *and* connotative differences between them. For instance, for the term pair '*persona in una situazione irregolare*' [person in an irregular situation] and '*illegal immigrant*', there is a semantic difference between 'illegal' and 'irregular', a morphosyntactic difference in the structure of the terms with the Italian term using [noun + prepositional phrase] compared with [adjective + noun], and a connotative difference, with '*persona*' [person] having a neutral connotation compared to the negative connotation often carried by the English term '*immigrant*'.

The manual allocation of values to each pair of terms according to this procedure does inevitably involve an element of subjectivity on the part of the researcher, who must make decisions about the level of similarity and difference between the terms involved. However, since the study is concerned with differences and changes between the two time periods, providing that the same researcher allocates the values across both time periods in a systematic manner following the above descriptions, there is an internal consistency to the methodology used here, which can produce informative data on terminological variation and convergence among the languages in question over time.

In addition, it should be noted at this stage that terminological distance between terms in different languages is of course perfectly normal and to be expected. Indeed, as pointed out in section 2.4, having no terminological distance between any terms in different languages would indicate an overly literal translation. For instance, the languages involved in this study evidently have different syntactic structures, and therefore the way in which different concepts are expressed as terms will inevitably be structured differently, even if semantic and connotative aspects are the same, for the purposes of an idiomatic translation. As a consequence, terminological distance between two terms does not necessarily mean that a term has been translated inaccurately.

Once they have been allocated, the terminological distance values are then weighted by multiplying them by the number of times a term pair appears in the corpus in order to ensure that these values are representative, thus producing a weighted terminological distance for each pair. The sum of these weighted terminological distances then needs to be divided by the total number of terms in that conceptual category to provide a value known as the interlingual variation index (IVI) (cf. Fernández-Silva & Kerremans, 2011). This is a mean value between 0 and 1, given to five decimal places, indicating the overall terminological distance between the two languages for that specific conceptual category. Table 10 below shows an example of this IVI calculation for Italian-English term pairs in the conceptual category UNACCOMPANIED MINOR in the 2015-2018 subcorpus.

Table 10: Example of the calculation of IVI for UNACCOMPANIED MINOR (IT-EN 2015-2018)

	IT	EN	Frequency	Terminological distance	Weighted terminological distance
UNACCOMPANIED MINOR	*Minore non accompagnato*	*Unaccompanied minor*	45	0	0
	Minore non accompagnato	*UAM*	1	0.5	0.5
	Minore non accompagnato	*Unaccompanied children*	1	0.5	0.5
	Minore rifugiato	*Refugee child*	7	0.5	3.5
	Totals		54		4.5
				Interlingual variation index	4.5/54 = 0.08333

Following this method, the lower the IVI for a pair of languages for a given conceptual category, the more terminologically similar those languages are for that concept. Since there is a significant disparity in the number of terms in different categories (e.g. in *EUMigrLaw Corpus* (1992-1998 EN-FR), the category REFUGEE contains 100 terms while the category SEASONAL WORKER has only 5), it is essential to calculate a weighted IVI for each category. This can be done by multiplying the IVI for each category by the appropriate weighting factor for each category expressed as a decimal. This weighting factor is determined based on the number of terms in the category as a percentage of the overall number of terms observed in the entire subcorpus. For example, the category MIGRANT in the 1992-1998 EN-FR section contains a total of 197 terms, out of an overall total of 818 terms in the entire 1992-1998 EN-FR subcorpus. Therefore, the weighting factor for this category is 197/818=0.24083. Hence, the weighted IVI for this particular category is: IVI 0.09137 x 0.24083 = 0.02200.

All of these weighted IVI values for each category can then be combined to produce a mean IVI value for the language pair in question, thus offering a single value that reflects the terminological distance between the two languages in question with regard to the migration-related terms being studied. This same process can then be repeated for the later time period for all three language pairs in order to provide mean IVI values for each one.

In order to verify the statistical significance of the results of this analysis, a Fisher's exact test can be carried out. In the same way as a Chi-squared test, the Fisher's exact test is used to determine whether the associations between different categorical variables are non-random, and is used in specific situations when the expected frequencies in a cross-tabulation are particularly small, as is the case here (Brezina, 2018). The Fisher's exact test can be conducted using the IBM SPSS Statistics software package (2022) by entering the individual weighted IVI values into a cross-tabulation. This will produce a p-value generated by the terminological distance analysis carried out. If the p-value is <.05, then the null hypothesis can be rejected and the result can be deemed to be statistically significant. The null hypothesis in the context of this investigation is that the differences identified in the analysis are purely the result of inherent variability in the corpus and cannot be attributed to any other factor.

This methodology will provide robust quantitative data to determine whether the interlingual variation index between the languages in question reduced between the two time periods (i.e. whether the languages have become terminologically more similar). Following the quantitative part of this stage indicating whether the terminological distances between the languages in question have changed, a qualitative analysis of the results can then be carried out to provide more insight into

the kind of changes which have occurred in the terminology of the three languages involved. While the quantitative analysis can provide a general answer to whether terminological convergence has occurred across the three languages in question, the qualitative analysis investigates where these changes have occurred and what form such changes have taken. The objective of the subsequent qualitative analysis is therefore to identify patterns of change across the languages in order to ascertain the direction in which any terminological convergence or divergence is moving.

4.2 Quantitative analysis

4.2.1 Intralingual terminological variation

Following the methodology described above, the *EUMigrLaw Corpus* was analysed to extract the key terms relevant to this study and place them into conceptual categories. Following this stage, a total of 15 conceptual categories were established across all languages for the earlier time period and 22 in the later time period. These categories are set out in table 11 below.

Table 11: Conceptual categories in the earlier and later time periods

Conceptual categories 1992-1998	**Conceptual categories 2015-2018**
ASYLUM SEEKER	ASYLUM SEEKER
CROSS-BORDER WORKER	CROSS-BORDER WORKER
DISPLACED PERSON	DISPLACED PERSON
IRREGULAR MIGRANT	IRREGULAR MIGRANT
LONG-TERM RESIDENT	LONG-TERM RESIDENT
MIGRANT	MIGRANT

MIGRANT WOMAN	**MIGRANT CHILD**
MIGRANT WORKER	MIGRANT WORKER
REFUGEE	**MOBILE WORKER**
RETURNEE	**OFFSHORE WORKER**
SEASONAL WORKER	**POSTED WORKER**
THIRD-COUNTRY NATIONAL	REFUGEE
UNACCOMPANIED MINOR	**RELOCATED PERSON**
UNION CITIZEN	**RESETTLED PERSON**
VICTIM OF TRAFFICKING IN HUMAN BEINGS	RETURNEE
	SEASONAL WORKER
	STATELESS PERSON
	THIRD-COUNTRY NATIONAL
	UNACCOMPANIED MINOR
	UNION CITIZEN
	VICTIM OF TRAFFICKING IN HUMAN BEINGS
	VULNERABLE PERSON

As shown in table 11 above, there was an increase in the overall number of concepts over the time period being studied with a number of new concepts emerging in the 2015-2018 time period that do not appear in the earlier one. The concepts only appearing in one of the two time periods are highlighted in bold in table 11. There is only one concept, MIGRANT WOMAN, that appears solely in the earlier time period, while the later time period contains eight concepts which do not appear in the earlier time period. This proliferation of new concepts can be explained by the development

in migration-related legislation within the EU over the time period in question. As described in chapter 3, the period between the two time periods witnessed the introduction of a series of new directives and systems, as well as large-scale treaty change, expanding the EU's competence in the domain of migration. Many of the new concepts appearing in the later time period (e.g. POSTED WORKER, RESETTLED PERSON, etc.) had yet to be established and legally defined in the earlier time period, and were introduced at a later point as the legislation in this area developed, hence their relatively recent emergence. This increase in terminology also reflects the fragmentation of the classification of migrants in response to the rising number of people arriving in Europe in a wide range of conditions and for many different reasons, triggering the need for 'terminological intervention' to describe and clarify them (Mariani, 2021, pp. 46).

The total number of terms identified for each language in each time period is shown in table 12 below, along with the mean, median and mode number of terms per concept to show the internal dynamic within each of the languages with regard to terminological variation, known as intralingual variation. The complete results can be consulted in appendix 1.1-1.6.

Table 12: Number of terms in each subcorpus by language

	Number of terms in 1992-1998 subcorpus	**Number of terms in 2015-2018 subcorpus**
English	**40** Mean: 2.7 terms per concept Median: 2 Mode: 1	**48** Mean: 2.2 terms per concept Median: 2 Mode: 1
French	**40** Mean: 2.7 terms per concept Median: 2 Mode: 1	**47** Mean: 2.1 terms per concept Median: 1 Mode: 1

Italian	**40** Mean: 2.7 terms per concept Median: 2 Mode: 1, 2	**47** Mean: 2.1 terms per concept Median: 2 Mode: 1

It is clear from the results shown in table 12 that there is widespread terminological variation within all three languages over both time periods. The results show a similar trend for English, French and Italian. In the earlier time period, a consistent total of 40 terms in each language gives a mean value of 2.7 terms used per concept - significantly higher than one term per concept. The lower median and mode values for each language indicate that there are a few conceptual categories containing a high number of terms, thus pushing up the mean values. Indeed, while certain concepts are consistently expressed in all languages with only a single term in this period, for instance SEASONAL WORKER and MIGRANT WOMAN, most concepts are expressed with several different terms in each language. Examples of concepts with at least two term variants in each language include: IRREGULAR MIGRANT, MIGRANT WORKER, THIRD-COUNTRY NATIONAL and UNION CITIZEN. To illustrate the spread of term variants used, the various terms expressing the concept UNION CITIZEN in the three languages in the 1992-1998 subcorpus are shown in figures 2-4 below.

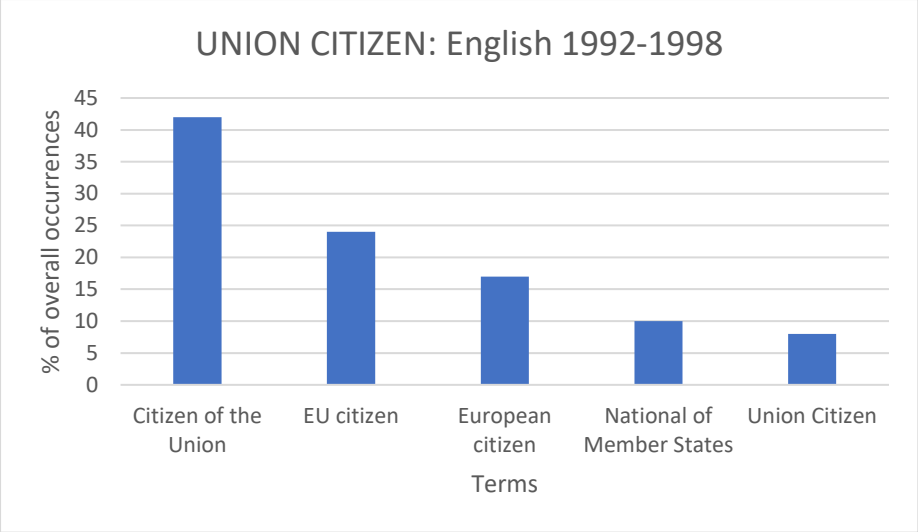


Figure 2: Terms used in English in 1992-1998 subcorpus for concept UNION CITIZEN

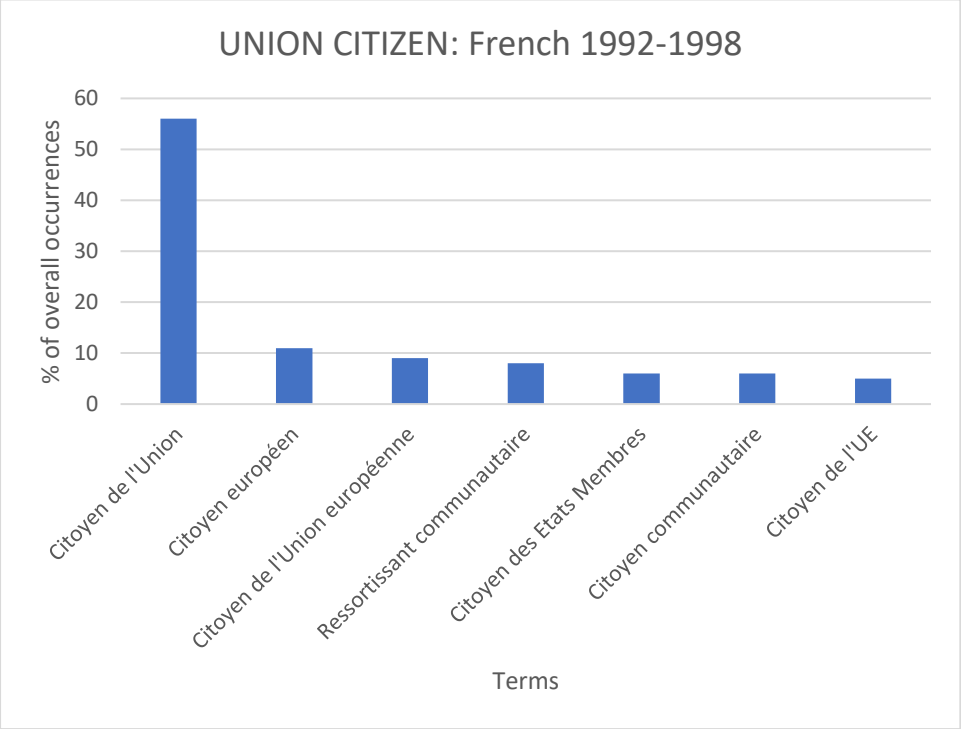


Figure 3: Terms used in French in 1992-1998 subcorpus for concept UNION CITIZEN

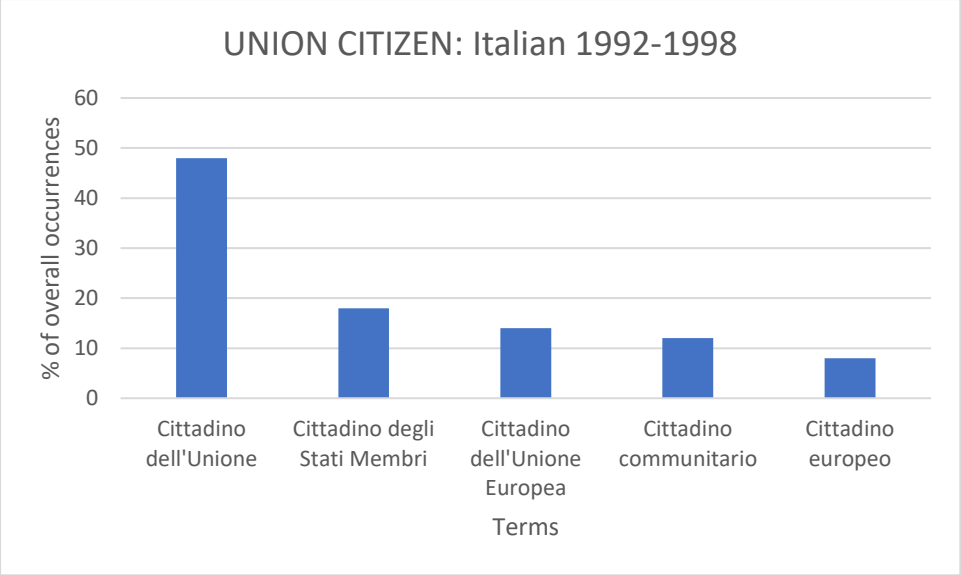


Figure 4: Terms used in Italian in 1992-1998 subcorpus for concept UNION CITIZEN

The above figures are illustrative of the level of variation and instability in terminology within each language for many of the concepts in the earlier subcorpus. These figures show the range of different ways of expressing the same concept in each language along with the frequency distribution of these various terms. The Italian and English subcorpora contain five term variants for this concept, while the French subcorpus contains seven. The extent of this intralingual variation is striking given the importance of consistency in legal texts, for the sake of both legal interpretation and public accessibility. Indeed, such variation is usually considered highly undesirable in legal language, especially in multilingual law (Biel & Koźbiał, 2020). Nonetheless, this finding is in line with previous research that identified significant intralingual synchronic terminological variation in various other domains of EU law, including criminal law (Peruzzo, 2013) and finance (Prieto Ramos & Morales Moreno, 2019).

The data in table 12 shows that the extent of this intralingual terminological variation reduces somewhat over the time period, with the average number of terms per

concept falling to 2.2 terms per concept in English and 2.1 terms in French and Italian. However, these results still indicate that more than one term is frequently used for the same concept in all three languages. Moreover, the median number of terms per concept for English and Italian remains at 2. Indeed, although fewer terms are used on average to express concepts in the later subcorpus, there are still numerous examples, such as VICTIM OF TRAFFICKING OF HUMAN BEINGS, MIGRANT WORKER and UNION CITIZEN, where at least two different term variants are used in all three languages. For purposes of comparison with figures 2-4, the various terms used for the concept UNION CITIZEN in the three languages in the 2015-2018 subcorpus are shown in figures 5-7 below.

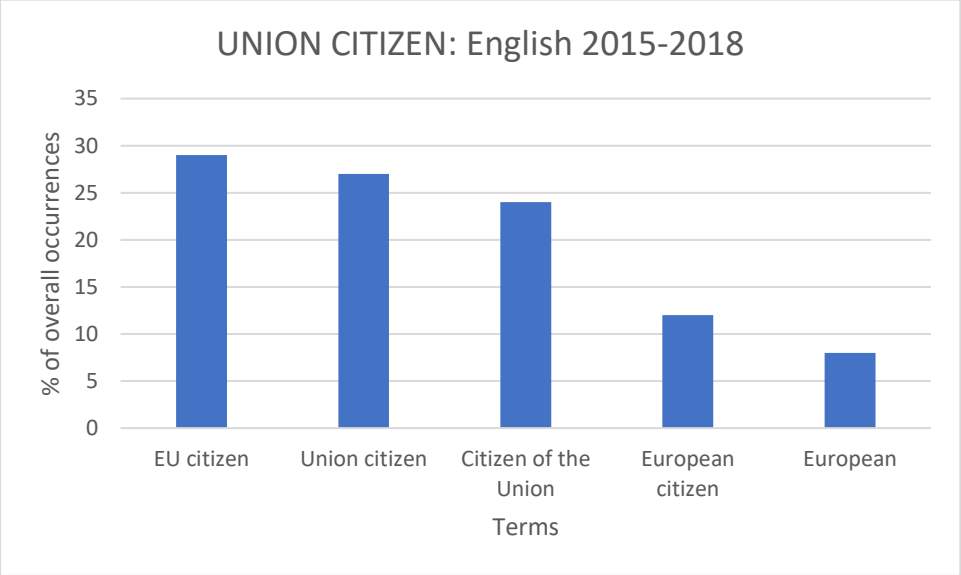


Figure 5: Terms used in English in 2015-2108 subcorpus for concept UNION CITIZEN

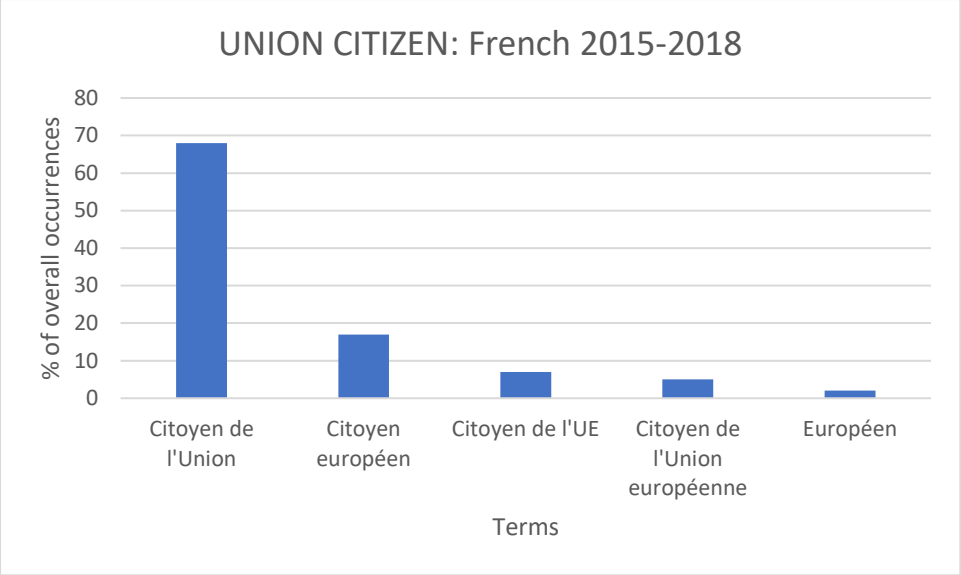


Figure 6: Terms used in French in 2015-2018 subcorpus for concept UNION CITIZEN

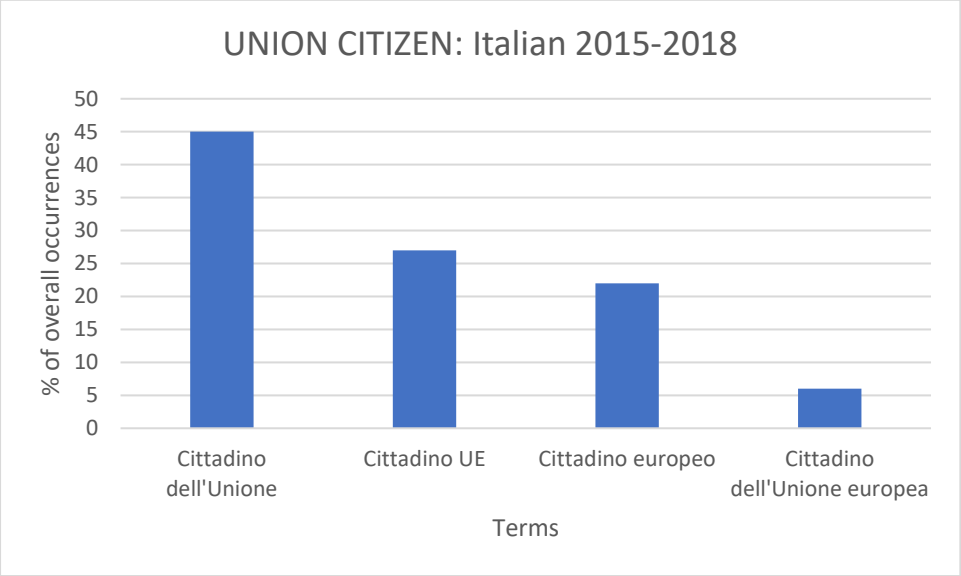


Figure 7: Terms used in Italian in 2015-2018 subcorpus for concept UNION CITIZEN

These figures show that, although the overall number of terms used to express the concept of UNION CITIZEN has reduced during the time period, there are still multiple terms which are used in all three languages. In English, although some of the terms differ from those used in the earlier subcorpus, there has been no reduction in the number of term variants for this specific concept. Moreover, the frequency distribution of these terms does not indicate that one term is becoming

predominant. The number of terms used in the other languages has decreased, from seven to five in French and from five to four in Italian. This indicates a slight increase in terminological stability for these languages, but only in French does there appear to be a clearly predominant term emerging (‘*citoyen de l’Union*’ [Citizen of the Union]). The persistence of numerous term variants in both languages demonstrates continued terminological variation for the concept UNION CITIZEN.

This slightly reduced but persistent intralingual terminological variation is characteristic of many of the concepts in the later subcorpus. Another example of this is evident with the concept of MIGRANT WORKER. There is a reduction in the number of terms in this category for each of the three languages: in English, the number of terms in this category falls from 4 to 3, in French, it reduces from 5 to 4, and in Italian, it drops from 6 to 4. Despite these consistent reductions, there remains multiple terms in this conceptual category. Moreover, as evident with the concept UNION CITIZEN above, no clearly predominant term has emerged for any of the three languages. Indeed, the most frequently occurring term in each language in this category fails to reach even 50% of the overall number of terms in that category, with ‘*economic migrant*’ accounting for 44% of the English category as a whole, ‘*migrante economico*’ [economic migrant] 41% of the Italian category and ‘*migrant économique*’ [economic migrant] just 35% of the French category.

The results in this section demonstrate that there is a great deal of variation and instability within the three languages for a significant number of the conceptual categories identified over both time periods. Although there is a moderate reduction in the number of terms per concept for each language, the data for the later subcorpus shows that the number of terms per concept remains significantly higher than one, indicating continued variability in expressing concepts in each language.

4.2.2 Interlingual terminological variation

Having observed the extent of terminological variation within the individual languages, the level of interlingual terminological variation can be analysed. This indicates how similar the terminology used across the three languages is for each concept and whether this has changed over the time period under investigation. As described in the methodology section above (4.1), the languages were analysed in a 'pair-wise' manner in order to obtain more fine-grained results on the interaction between the languages in question and to ensure that there is no *a priori* assumption regarding the translation direction. In this analysis, English terms are compared to French, French to Italian, and Italian to English. This circular approach ensures that each language is given equal footing and produces a list of term pairs for each conceptual category for each language pair. An example of the term pair results for the concept DISPLACED PERSON in the earlier time period is provided in table 13 below, showing the frequency of occurrence and terminological distance score (0, 0.5 or 1) of each pair, along with the calculations of weighted terminological distance and ultimate interlingual variation index for that particular conceptual category for this language combination.

Table 13: Term pairs for DISPLACED PERSON in earlier subcorpus

English	**French**	**Frequency**	**Terminological distance**	**Weighted terminological distance**
Displaced person	*Personne déplacée* [Displaced person]	16	0	0
Displaced person	*Déraciné* [Uprooted]	1	1	1
Displaced person	*Personne expatriée* [Expatriated person]	1	0.5	0.5
Totals		18		1.5
Interlingual variation index				1.5/18 = 0.08333
French	**Italian**	**Frequency**	**Terminological distance**	**Weighted terminological distance**
Personne déplacée [Displaced person]	*Sfollato* [Displaced]	12	0.5	6
Personne déplacée [Displaced person]	*Profugo* [Refugee][9]	2	0.5	1
Personne déplacée [Displaced person]	*Persona sfollata* [Displaced person]	1	0	0
Personne déplacée [Displaced person]	*Rifugiato* [Refugee]	1	1	1
Personne expatriée [Expatriated person]	*Sfollato* [Displaced]	1	1	1
Déraciné [Uprooted]	*Sfollato* [Displaced]	1	0.5	0.5

[9] See above for more detailed discussion of *profugo.*

Totals		18		9.5
Interlingual variation index				9.5/18 = 0.52778
Italian	**English**	**Frequency**	**Terminological distance**	**Weighted terminological distance**
Sfollato [Displaced]	*Displaced person*	14	0.5	7
Persona sfollata [Displaced person]	*Displaced person*	1	0	0
Totals		15		7
Interlingual variation index				7/15 = 0.46667

In accordance with the methodological steps set out above, a series of tables for each language pair and both time periods can be created, as per table 13 above. These complete results tables can be consulted in the appendix. Using this data, the weighted IVI values for each conceptual category can be calculated, as described in section 4.1 above. For instance, taking the example of the English>French category in table 13 above, the overall number of term occurrences in this category for this time period is 18. The total number of term occurrences for the English>French part of the earlier subcorpus is 818, meaning that the category DISPLACED PERSON thus represents 2.2% of the total term occurrences. In order to calculate the weighted IVI value, the IVI for the category (0.08333) therefore needs to be multiplied by (0.022), which gives a weighted IVI value for this category of 0.00183.

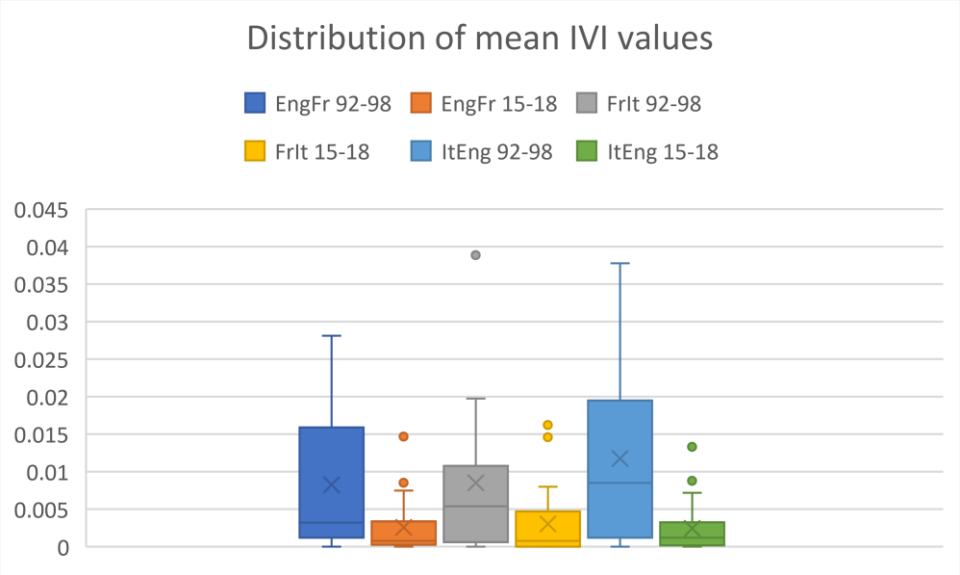


Figure 8: Distribution of mean IVI values by language pair and time period

Figure 8 above shows the distribution of mean IVI values for each category and each language pair over both time periods. It is clear that there is a much wider distribution of values in the earlier time period for all language pairs compared to the distribution of values in the later time period. This shows that terminological distance between the languages in question was substantially wider in the earlier time period than in the later one.

With the aid of these weighted IVI values, the mean IVI for each language pair and both time periods can be calculated, and are set out in table 14 below, along with the percentage change for each language pair over the time period.

Table 14: Mean IVI values

	Mean IVI 1992-1998	**Mean IVI 2015-2018**	**% change between time periods**
English > French	0.00820	0.00258	**68.69% decrease**
French > Italian	0.00867	0.00304	**64.94% decrease**
Italian > English	0.01177	0.00255	**79.10% decrease**

The results in table 14 above show a consistent trend across the three language pairs. The mean IVI values for each language pair decrease by between 64.94% (French>Italian) and 79.10% (Italian>English). This suggests a marked fall in interlingual terminological variation between the time periods and a level of terminological convergence across the three languages. Although the degree of the terminological convergence observed for all three language pairs is relatively comparable, it is not entirely uniform and a number of differences can be observed. The largest terminological distance identified in the earlier time frame is between Italian and English, indicated by the largest IVI value of 0.01177. However, the Italian-English language pair then becomes the most terminologically similar in the later time frame, with an IVI value of 0.00255. This equates to a decrease in IVI value of 79.10% - the largest decrease among the three language pairs. The language pair with the lowest IVI in the earlier time frame, English-French (0.00820), records a smaller percentage decrease in its IVI value (68.69%), leaving it with a slightly higher IVI value than the Italian-English pair for the later time period (0.00258). The IVI values for the French-Italian language pair fall in between the other two language pairs in the earlier time period but this language pair has the

smallest percentage decrease in IVI value over the time period, leaving it with the highest IVI value in the later time period (0.00304). Therefore, although the change in terminological distance across all the language pairs moves in the same direction indicating convergence in each case, the magnitude of this convergence differs depending on the specific language pair.

As mentioned in the methodology above, the Fisher's Exact statistical significance test can be conducted on these results to verify that the differences detected between the IVI values over the time period are not purely due to chance. The aim of this test is to ascertain whether it is possible to reject the null hypothesis that the differences identified by the analysis are simply due to random variability in the corpus. The IVI values for each concept in each language pair were therefore placed in a cross-tabulation using the IBM SPSS Statistics software package in order to determine the p-values for each language pair. These p-values are shown in table 15 below:

Table 15: p-values derived from Fisher's Exact test for each language pair

	p-value
English > French	0.006
French > Italian	0.067
Italian > English	0.031

According to the p-values obtained from the Fisher's exact test, the results for the English>French and Italian>English language pairs can be considered statistically significant since they fall below the cut-off point of $p<.05$, and the null hypothesis can be satisfactorily rejected for these pairs. It can therefore be argued that the

differences between the IVI values for these language pairs across the time period in question are statistically significant and are due to some reason other than chance. However, the p-value of 0.067 for the French>Italian language pair means that the IVI values for this pair fall below the threshold for rejecting the null hypothesis, and in statistical terms, there remains the possibility for this pair that the change in IVI values is the result of random chance.

This statistically significant decrease in the IVI values for the pairs English>French and Italian>English consequently indicates that there has been significant terminological convergence across these pairs, at least with regard to the migration-related terms being investigated here. This conclusion can be reached from the quantitative data above because the IVI value is an indication of the mean terminological distance between two languages with regard to the set of concepts under investigation in the study. A decrease in the IVI therefore indicates a drop in the mean terminological distance observed for the collection of migration-related terms being examined. In other words, there has been a marked increase in the semantic, morphosyntactic and/or connotational similarities between the terms being used in the different languages. The magnitude of the percentage decreases in the IVI values for these pairs (68.69% and 79.10%) also demonstrates the extent of convergence that has taken place between the language pairs.

However, it is not possible to make the same observation of terminological convergence for the French>Italian language pair, since the IVI results for this pair do not prove to be conclusively statistically significant. Indeed, the reduction in the mean IVI for this language pair is not as dramatic as for the other two, and the mean IVI value for the later time period remains higher than the other two pairs. Consequently, taken together, these results tend to point to English as the common

denominator in the terminological convergence that has occurred and could therefore constitute the point towards which French and Italian have converged over time. This hypothesis is investigated further in the qualitative analysis in section 4.3 below.

This quantitative data therefore provides robust, empirical evidence of the considerable interlingual convergence that has occurred for the set of migration-related terms examined, as well as the differences in the extent of this convergence across the three language pairs. While the quantitative results are able to clearly demonstrate the direction of terminological development tending heavily towards convergence, they are unable to provide an insight on the influence of the languages on one another. As a result, a complementary qualitative analysis is necessary to examine the terms on an individual basis in order to gain a more detailed insight into the kind of terminological changes which occur across the time period.

4.3 Qualitative analysis

4.3.1 General observations

Firstly, a number of general observations can be made when examining the individual concepts and terms in more fine-grain detail. It is important to underline at this stage the significant level of imprecision and discrepancies which are observed both within the individual languages and across the three languages being studied. Closer analysis of the results produced shows that, especially in the earlier subcorpus, terms with markedly different meanings are frequently used interchangeably, leading to clear discrepancies between different languages. In the past, such discrepancies have led to legal uncertainty on a number of occasions,

making it necessary for courts to consult different language versions in order to establish the meaning of legislation beyond reasonable doubt (e.g. ‘Srl CILFIT v Ministry of Health’ (1982); ‘Allsports Ltd v UK Office of Fair Trading’ (2003); ‘JJB Sports PLC v UK Office of Fair Trading’ (2004)). A clear example of such discrepancies in this study can be seen in the conceptual categories DISPLACED PERSON, REFUGEE and MIGRANT. Indeed, where the English and French texts use the terms ‘*displaced person*’ and ‘*personne déplacée*’ [displaced person], the Italian texts occasionally use ‘*rifugiato*’ or ‘*profugo*’ [refugee], as evidenced in table 13 above. Likewise, the Italian texts occasionally use ‘*migrante*’ [migrant], where the English and French texts use ‘refugee’ and ‘*réfugié*’ [refugee]. The legal definitions of these terms are strictly delineated and it is therefore surprising to see such widespread inconsistency in their use between the languages concerned. A further example of such inconsistency is evident in the conceptual category VICTIM OF TRAFFICKING IN HUMAN BEINGS. In the Italian texts in the earlier subcorpus, the terms ‘*vittima del traffico*’ [victim of smuggling] and ‘*vittima della tratta*’ [victim of trafficking] are used interchangeably. Once again, there is a clear legal difference between these terms. Trafficking in human beings ‘involves the recruitment, movement or harbouring of people for the purpose of exploitation’ while migrant smuggling entails ‘assisting migrants to enter or stay in a country illegally’ (UNODC, 2018). The apparently indiscriminate use of these terms in Italian is thus highly problematic from a legal perspective because victims of trafficking are entitled to a special set of protections under international and European law, which would not be available to them if they were not correctly identified as such (Pace and Severace, 2016, pp. 69). These inconsistencies and discrepancies help in part to explain the

extent of terminological inconsistency identified by the quantitative analysis above, especially in the earlier time period.

It should also be noted that the total number of terms in each category is not always the same for each language. There are two main reasons for this discrepancy. Firstly, as discussed above, terminological imprecision and errors in the texts means that terms used in different languages occasionally have to be classified in different conceptual categories. For example, a text in the earlier subcorpus refers to '*migrant workers*' in the English version, while the French and Italian versions refer to '*étrangers*' [foreigners] and '*stranieri*' [foreigners] respectively. This means that the English term is classified in the MIGRANT WORKER conceptual category, while the French and Italian terms belong in the THIRD-COUNTRY NATIONAL category. Likewise, the Italian-English term pair '*rifugiato* [refugee]-*displaced person*' is placed in the REFUGEE conceptual category, while the French-Italian term pair '*personne déplacée* [displaced person] - *rifugiato* [refugee]' is placed in the DISPLACED PERSON category. Secondly, the different language versions of the texts occasionally take different discursive approaches. This means that a term may be repeated in one language version, while another uses an elliptical construction to avoid repetition. An example of this situation with the concept THIRD-COUNTRY NATIONAL in the English-French language pair in the earlier subcorpus can be seen here:

> English: *For a third-country national entering the EU without a visa, there are two possible scenarios*.
>
> French: *Dans le deuxième cas (entrée sans visa), deux situations peuvent se présenter*.

> [In the second case (entering without a visa) two situations are presented].
>
> Opinion of the Economic and Social Committee on the 'Proposal for a Council Directive on the right of third-country nationals to travel in the Community'
>
> *Official Journal C 153, 28/05/1996 P. 0038*

Occurrences such as the above example also contribute to the occasional differences between the languages regarding the overall number of terms for each conceptual category.

In addition, it is interesting to note that the terminological variation within some of the conceptual categories which only appear in the later subcorpus is often lower than for concepts which appear in both and have a longer history of usage in EU law. Indeed, a number of newer concepts have lower IVI values across all three language pairs compared to the average IVI, namely MOBILE WORKER, OFFSHORE WORKER, POSTED WORKER, RELOCATED PERSON and VULNERABLE PERSON. This finding is contrary to the general principle that terms for new concepts tend to be more unstable and susceptible to variation, whereby the lack of terminological standardisation leads to a proliferation of synonyms and variants (e.g. Bertaccini, Massari & Castagnoli, 2010). It has been frequently observed that when terminology is initially used for a given concept, there tends to be greater variation and instability, which Freixa refers to as ‘chronological variation’ where various forms co-exist ‘until certain forms achieve a priority level over others’ (2006, pp. 56). It is possible that these terms emerged in the gap between the time periods covered by the corpus, and that by the time of the later subcorpus these terms had already become standardised, thus the initial stage of variation and instability was not captured by the corpus data. However, this would fail to explain why there remains

significant terminological variation for concepts that have been in use for much longer time periods.

Indeed, this finding can be more satisfactorily explained by referring back to the EU's unique approach to multilingual term creation. If new EU legislation introduces a new concept, a specific term in each official language must be established to express such a concept, especially if this concept is not known or developed beforehand in national jurisdictions. As described in section 2.2, terms tend to be created in the dominating languages of English, French and German in a process known as 'primary-term creation', before then being translated into the other languages (Fischer, 2010, pp. 28). The process of term creation within the EU system is therefore a wholly intra-conceptual activity, whether arrived at through term creation or translation (ibid., 2010). The concept is defined by the subject experts, politicians and civil servants involved in the legislative drafting process in collaboration with terminologists and translators in advance of the term creation stage (Čavoški, 2020). These newly introduced, often EU-specific, concepts therefore have relatively low terminological variation as a result of the highly controlled term creation process involved. In contrast, the terms with the highest levels of variation and instability tend to be those for concepts which were not created within the EU conceptual system but which have developed separately and over longer periods within the languages being studied (e.g. IRREGULAR MIGRANT).

4.3.2 Influence of English terminology

Having noted the differences between the general spread of concepts and terms, the content of the individual conceptual categories can now be analysed. Although it is

important to remember that it is very uncommon for contact-induced language change in any setting to have only one identifiable cause, an examination of the development of the individual conceptual categories provides evidence that the translation of English terminology into French and/or Italian has an influence on the terms used in those languages. This evidence provides support for the hypothesis emerging from the quantitative analysis, which indicated that French and Italian terminology was converging towards English terminology. A qualitative analysis of the data indicates that there are a significant number of concepts for which English seems to have exerted some influence through translation over French and/or Italian resulting in substantial semantic, morphosyntactic and/or connotational changes in the terminology used in those languages. In total, there is evidence of influence from English on Italian terminology in six conceptual categories, namely: IRREGULAR MIGRANT, LONG-TERM RESIDENT, MIGRANT WORKER, OFFSHORE WORKER, REFUGEE and RETURNEE. There is evidence of influence from English on French terminology in three conceptual categories: LONG-TERM RESIDENT, MIGRANT WORKER and OFFSHORE WORKER. This section provides an illustrative selection of these concepts to exemplify this hypothesis.

A first example of a concept which shows evidence of influence from English is LONG-TERM RESIDENT, which refers to anyone with long-term resident status under the Long-Term Residents Directive (Council Directive 2003/109/EC, 2003). In both subcorpora, the only term used in English for this concept is '*long-term resident*'. However, the terms used in both French and Italian for this concept differ between the two subcorpora. The French term used on all occasions in the earlier time period is '*(personne) durablement installé*' [permanently settled (person)] and in Italian it is '*persona stabilitasi a titolo duraturo*' [person settled on a long-term basis].

In the later subcorpus, both these languages drop this terminology entirely, shifting to '*résident de longue durée*' [long-term resident] and '*soggiornante di lungo periodo'* [long-term resident] in all cases, which both closely reflect the terminology used in English. For this concept, the French and Italian terms move closer to English both semantically and morpho-syntactically and the Italian term replaces the word '*persona*' [person] with the more specific label '*soggiornante*' [resident].

Consequently, there is evidence that the terminological shift in French and Italian for the concept of LONG-TERM RESIDENT is influenced by the translation of the terminology used in the English texts. In addition, the terms used in French and Italian are also the terms now recommended by the EMN glossary. The IATE database specifically labels the French formulation '*durablement installé*' as 'deprecated', meaning it is acknowledged as having been used previously, but is now discouraged. It is therefore likely that instructions provided to translators regarding terminology have also played a role in this shift, although these instructions may well have themselves been influenced by English. The potential influence of these institutional terminology recommendations is examined in more depth in the discussion section below.

Another concept which undergoes considerable internal change during the time period, which also appears to be in part the result of influence from English, is the category MIGRANT WORKER. A migrant worker can be defined simply as a person engaged in 'a remunerated activity in a state of which they are not nationals' (Directorate-General for Migration and Home Affairs, 2020). In the earlier subcorpus, the most frequently occurring term for this concept in English, French and Italian is '*migrant worker*' (62% of cases), '*travailleur migrant*' [migrant worker] (72% of cases) and '*lavoratore migrante*' [migrant worker] (76% of cases) respectively. In English,

the second most popular alternative term for this concept is '*foreign worker*' (15% of cases). The equivalent French and Italian terms for '*foreign worker*', '*travailleur étranger*' (9% of cases) and '*lavoratore straniero*' (8% of cases), do not appear as frequently. By the later period, '*foreign worker*' increases significantly as a proportion of the conceptual category in English (25% of cases), while '*travailleur étranger*' and '*lavoratore straniero*' occur with almost the same frequency as '*travailleur migrant*' and '*lavoratore migrante*' (24% of cases in both). The trend for this concept moves towards an increased use of the adjective '*foreign*' (and its French and Italian equivalents) over time, which is led by English terminology and followed by French and Italian. This marks a significant connotational shift in the terminology used for this concept. While the connotations of the adjective '*migrant*' in this context are relatively neutral, the adjective '*foreign*' and its French and Italian equivalents have much more negative connotations, evoking undertones of otherness and alienation. Once again, it appears that the influence of English through translation can override prescriptive efforts of terminological control since this trend runs contrary to terminology recommendations provided by the EMN glossary and the IATE database, which advocate the use of '*migrant worker*' and its French and Italian equivalents.

Furthermore, it should be noted that in this category MIGRANT WORKER, all three languages have the same new predominant term in the later time period, namely '*economic migrant*' (50% of cases), '*migrant économique*' (41% of cases) and

'*migrante economico*' (41% of cases)[10]. Once more, there appears to be influence from English, in this case specifically on Italian, with regard to this specific term. Indeed, in the earlier time period, this concept appears to have undergone very little terminologisation in Italian, since it is expressed in various prolix manners, such as '*cittadini di paesi terzi che cercano di entrare […] per motivi economici*' [citizens of third countries who try to enter […] for economic reasons] and '*(persone) che emigrano per motivi economici'* [(people) who emigrate for economic reasons]. However, by this point, English has an established term, 'economic migrant', to express this concept, which is already used in 9% of occasions in the earlier subcorpus. However, by the later subcorpus, Italian terminology has shifted to mirror the English terminology with the term '*migrante economico*' being used on 41% of occasions.

In this case, for the conceptual category MIGRANT WORKER, translation from English appears to have had an influence on the changes occurring within it over the time period. Indeed, the proportional increase in the use of '*travailleur étranger'* [foreign worker] and '*lavoratore straniero*' [foreign worker] comes in the wake of higher use of the English term '*foreign worker*' in the earlier time period, while the appearance of the Italian term '*migrante economico*' [economic migrant] in this same category also reflects the usage of the English term '*economic migrant*' from the earlier time period.

[10] While the EMN glossary identifies a slight difference in its definitions of '*economic migrant*' and '*migrant worker*', whereby the former is 'not in any way related to the refugee definition', the ILO points out that they are most often used as synonyms, and that the term "*migrant worker*", as defined in the *International Convention on the Rights of all Migrant Workers and Members of Their Families* (1990), Art. 2.1), is a more neutral term and should be preferred (2020, pp. 14). For this reason, these terms were placed in the same conceptual category for the purposes of this research.

As noted above and explored in more detail below, there are more instances where the influence of English is apparent on Italian terminology than on French. This is the case, for instance, for the concept RETURNEE, defined in the European Migration Glossary as ‘a person going from a host country back to a country of origin […] usually after spending a significant period of time in the host country whether voluntary or forced, assisted or spontaneous’ (Directorate-General for Migration and Home Affairs, 2020). The only term used in English to express this concept in both subcorpora is ‘*returnee*’. In the earlier subcorpus, this concept is expressed most frequently in French by ‘*candidat au retour*’ [candidate for return] (50% of cases) and in Italian by ‘*persona che intende rientrare*’ [person planning to return] (66% of cases). Both languages therefore opt for the use of a multi-word term with a preposition or relative clause in contrast to the single word term used in English. However, the morphosyntax of the Italian term shifts entirely in the later subcorpus, as the concept RETURNEE is expressed in the majority of cases (88%) by the term ‘*rimpatriato*’ [repatriate]. The relative clause structure has been abandoned in Italian and replaced by a single-word term, mirroring the English term structure for this concept. This is despite the fact that the Italian term recommended by the EMN glossary and the IATE database retains the previous morphosyntactic structure: ‘*persona che compie ritorno*’ [person undertaking a return]. This therefore suggests that the terminological choice of the translator is being influenced by something other than terminology prescribed by the EU’s own terminology services. Indeed, as argued above, we can assume that most primary term creation by this point occurs in English and secondary terms are then derived through translation, and that this contact with English in translation may well therefore be involved in the shift in the morphosyntax of this term.

Studies have shown that translation-induced language change can affect not only lexical features, but also the morphosyntactic structure of a target language (e.g. Amouzadeh & House, 2010; Bisiada, 2016), whereby the word-formation or sentence-formation conventions of a target language are modified as a result of contact with another language through translation. In this instance, the structure of the Italian term for the concept RETURNEE has grown very similar to the structure of the English term with no trace whatsoever of the previous Italian term structure remaining in the later subcorpus despite prescriptions to the contrary in the EU's own terminology database and glossary. This would suggest that these Italian texts are produced through translation of an English source text, which thus influences the terminology used in the Italian text. Brannan (forthcoming, pp. 29) suggests the origin of this usage in Italian is the EU's Return Directive of 2008, where the English term '*returnee*' 'proved awkward to translate into other languages', further reinforcing the notion that the Italian term '*rimpatriato*' was arrived at through the translation of the English term.

The knock-on effect in this case is the removal of any notion of agency from those people in this category. The Italian terms in the earlier subcorpus and in the institutional recommendations, such as '*persona che intende rientrare*' [person planning to return] and '*persona che compie ritorno*' [person undertaking a return], give the sense that the people involved have a level of control over the action of returning. However, the new term '*rimpatriato*' [repatriate], which echoes the English term, suggests that they have no command whatsoever over the action that is being done to them. Interestingly, the term '*rimpatriato*' has also become increasingly commonly used domestically since its appearance in EU law (Brannan, forthcoming),

pointing to the influence of EU legal terminology and, by extension, the significant potential impact of translation-induced language change at this level.

Moreover, it is interesting to note that, in this case, French terminology in the later subcorpus still has a large degree of variation (six different French terms are used for the concept RETURNEE in total) but in large part retains the multi-word term structure with '*personne soumise à un retour*' [person subject to a return] (81% of cases) being the most frequently occurring term. Although there has been some internal shift (the predominant French term in the earlier time period being '*candidat au retour'* [candidate for return] (50% of cases)), the multi-word term structure remains. Unlike the Italian, the French morphosyntactic structure for this term does not appear to be influenced by English. It is also important to note that according to the EMN glossary, French, unlike English and Italian, makes a distinction between a forced and a voluntary return, using the terms '*personne éloignée*' [distanced person] and '*personne retournée*' [returned person] respectively. This further strengthens the argument that where concepts have initially developed independently of the EU system, there tends to be greater terminological variation and differences between the languages at both a conceptual and terminological level.

4.3.3 The case of IRREGULAR MIGRANT

One conceptual category which merits particular attention is that of IRREGULAR MIGRANT. This is both because it is a central concept in EU migration law and because terminology used to express it shows signs of influence from both English and French through translation. It is the only concept from those studied here to demonstrate this pattern and highlights the complex network of influences involved in

shaping and evolving terminology in this context. This concept, defined in the EU context as a 'third-country national' in an EU Member State who does not fulfil 'the conditions for entry, stay or residence' in that State (Directorate-General for Migration and Home Affairs, 2020), has long been the subject of debate and controversy. Indeed, the IOM notes that 'a universally accepted definition of irregular migration does not exist' (IOM, 2022). As discussed in section 2.3, use of the term 'illegal' in this context has long been considered by many as both inappropriate and legally incorrect (Guia, 2016), but disagreement continues regarding the most appropriate term to use for this concept. Various suggestions have been made in English for the best alternative term, including 'illegalized immigrant' (Bauder, 2014), 'non-documented migrant' (United Nations General Assembly, 1975), 'migrant without papers' (Council of Europe Parliamentary Assembly, 2006) and 'irregular migrant' (European Parliament, 2009). This ongoing terminological uncertainty and instability is embodied in the terminological variation still apparent in all three languages being studied.

Indeed, in the 1992-1998 subcorpus, the most frequently occurring terms for this concept for English, French and Italian respectively are '*illegal immigrant*' (76% of cases), '*immigré en situation irrelulière*' [immigrant in an irregular situation] (49% of cases) and '*immigrato in situazione irregolare*' [immigrant in an irregular situation] (46% of cases). The adjective '*irregular*' predominates in the earlier time period in French terminology (used in 70% of cases) and, to a lesser extent, in Italian terminology (used in 56% of cases) but is only very marginal in English (only used in 16% of cases). However, in the later part of the subcorpus, '*irregular*' becomes the leading adjective used in the expression of this concept in English too, and becomes even more dominant in French and Italian. It is used in 85% of cases in English, 91%

in Italian and 100% in French. The data consequently shows that French was earlier to adopt the adjective in this term, and that English especially adopted it much later. It is likely that the use of the term '*irregulier*' for this concept in French EU texts influenced the switch in English from '*illegal*' to '*irregular*'. This hypothesis that French can still exert an influence through translation is supported by other research in this area that highlights the conventionalisation of the term '*rehlament*' [regulation] in the Ukrainian version of EU legislation[11], arguing that this term is more likely to have been derived from a French source than an English one (Pavliuk, 2023).

As well as the adjective used in this term, it is also interesting to note that there is an increasing trend in French to use the term '*personne*' [person] in place of '*migrant*', which reflects other findings from corpus-based research that indicate a gradual move in EU texts towards more 'humanising' terminology (Mariani, 2018, pp. 160), although this is not (yet) mirrored by the English or Italian terms for this particular concept.

Although the lexical features of the terminology in this category seem to have been influenced by French, English terminology has nonetheless appeared to have exerted influence through translation on Italian morphosyntax. Indeed, the structure of the most frequently occurring Italian term in the earlier subcorpus closely resembles the structure of the most common French term over that time period: '*immigrato in situazione irregolare*' and '*immigré en situation irrelière*' [immigrant in an irregular situation]. Both languages use a prepositional phrase to modify the noun. In this construction, it is the situation that is described as irregular. By contrast, the structure of the English term differs. Although the predominant adjective changes

[11] Ukraine became an EU accession country in 2022.

from '*illegal*' to '*irregular*' over the time period, the adjective always refers directly to the immigrant or migrant in both time periods: either '*illegal immigrant*' in the earlier subcorpus or '*irregular migrant*' in the later one. An analysis of the results for this concept in the later subcorpus shows that the Italian terminological structure has shifted from the 'French' structure with a prepositional phrase to the 'English' structure. Indeed, the most frequent Italian term in the later subcorpus is '*migrante irregolare*' [irregular migrant] (82% of cases), with the adjective directly modifying the noun '*migrante*'.

While this may appear a minor morphosyntactic change, its effect should not be understated. Indeed, there is a subtle but very significant difference between describing a *situation* which a person happens to find themselves in as irregular and describing a *person* as irregular. In this case, the focus moves away from the *transgression* and shifts to the *transgressor*. It therefore appears here that the Italian morphosyntax has been altered under the influence of English through translation, which has led to a fundamental modification in the way in which the Italian legal texts address those who, for whatever reason, have failed to comply with the conditions of entry, stay or residence in force. Conversely, there is no change in the morphosyntax of the French terms for this concept over the time period. Indeed, in the later subcorpus, almost all the French terms retain the prepositional phrase structure, with the adjective '*irregulier*' applying to either the '*situation*' or the '*séjour*' [stay].

A qualitative analysis of the terms used for the concept IRREGULAR MIGRANT therefore indicates that influence has been exerted in different ways by both French and English in translation. This finding is supported by other corpus-based research into multilingual EU legal texts, which concludes that 'Italian Eurolect lexical choices mirror lexical variants adopted in master texts (basically in English and French) and,

consequently, they may also be interpreted as contact-induced phenomena' (Mori, 2018b, pp. 208). This is possible due to the unusual translation process employed by the EU system, as examined in section 2.2, in which 'there is no one target text which has an unambiguous relation to one specific 'original'' (Dollerup, 2004, pp. 197). Indeed, it is conceivable that at different points both English and French served as *de facto* source languages for certain texts, parts of texts or specific terms, and that they therefore both played a role in shaping the terminology used for the concept IRREGULAR MIGRANT. The shifting terminological profiles of the three languages for this concept serves to highlight the complex interplay at work between these languages. Indeed, the lexical features of the most common term used in English are influenced by French terminology through its adoption of the adjective '*irregular*', while the morphosyntactic structure of the most frequently used Italian term is influenced by the English term structure, with the noun '*migrante*' directly modified by the adjective '*irregolare*'.

4.4 Summary of results

The results of the corpus analysis are able to provide robust answers to the research questions raised by this stage of the corpus analysis. Firstly, the findings indicate very widespread intralingual terminological variation (variation within the individual languages) among migration-related terms in all three languages. The extent of this intralingual variation decreases somewhat over the time period, indicated by lower numbers of term variants per concept. This reduction in terminological variation within each language is undoubtedly a positive sign in the interests of legal clarity. However, the variation still present remains significant. This is especially striking

given the legally binding nature of the texts in question and the recognition that such synonymy ‘works to the detriment of clarity’ (Strandvik, 2012, pp. 39). Indeed, even in the later subcorpus, many concepts are still expressed by multiple term variants in all three languages, highlighting the persistent nature of this variation.

The quantitative analysis of the corpus findings also clearly indicates considerable interlingual variation in the earlier subcorpus. The results demonstrate terminological convergence between all three language pairs in the later subcorpus, indicated by the dramatic decrease in IVI values for each language pair. The Fisher’s Exact test demonstrates the statistical significance of these findings for the language pairs English>French and Italian>English, but not for French>Italian, for which the null hypothesis cannot be rejected. This is due to the fact that the convergence between French and Italian is not as marked as the convergence of the other two pairs, with the mean IVI for the French>Italian language pair remaining higher than for the other two language pairs. This suggests terminological convergence in the direction of English, with terminology in French and Italian moving to become more similar to English terminology.

A complementary in-depth qualitative analysis of the corpus findings provides substantial evidence for the influence of English on Italian and, to a lesser extent, French terminology through translation. Indeed, French and Italian terminology for a number of concepts, a selection of which are examined above, shows signs of influence from English, manifested in lexical, morphosyntactic and connotational changes to the French and Italian terminological profiles for those concepts over the time period. Although this study focuses specifically on migration terminology, there is no reason to assume that this phenomenon should be limited solely to this particular area of EU competence. Since translators depart increasingly often from

primary terms in English, this ‘inevitably influences their choice of target language designation of the same concept’ (Fischer, 2010, pp. 29). Indeed, there is a tendency among translators in the EU to resort to ‘calques’ when transposing terms from the language in which it was initially created into other languages in order to imitate the lexical structure of the terms, while paying little heed to the conceptual aspects of the terms (Kardos, 2023). Such literal transposition of terms across language boundaries is also bound to contribute to the convergence identified in this part of the study. As shown in sections 4.3.2 and 4.3.3 above, this convergence also results, on occasion, in an alteration of the connotations of terms used, often in a more negative direction, such as with the increased use of ‘*travailleur étranger*’ [foreign worker] in French, and ‘*lavoratore straniero*’ [foreign worker] and ‘*migrante irregolare*’ [irregular migrant] in Italian, or in a way that removes any agency from the migrants in question, such as with the emergence of the Italian term ‘*rimpatriato*’ [repatriate].

In addition, the qualitative analysis indicates that there is nonetheless a high degree of complexity involved in the interaction between the languages being studied, since there is also evidence of the influence of French on English terminology for the concept of IRREGULAR MIGRANT, and a greater susceptibility of Italian to influence through translation compared to French. Indeed, the influence from English in translation does not exert itself in a uniform way on both languages or for all concepts identified. This can be explained by the different statuses and linguistic cultures of the languages under investigation as well as the fact that there are almost always multiple internal and external factors which bring about language change, and those factors likely differ from one term to another (Thomason, 2006). These factors are examined in more depth in the discussion section below. In any case, the findings of this part of the corpus analysis set out in this research demonstrate that

among the undoubtedly complex network of influences, the impact of language contact through translation is present and able to exert sufficient influence to contribute to changing the way in which certain migration-related concepts are expressed as terms in the languages studied here. In sum, the results of this part of the research project provide strong evidence of the potential of language contact through translation to lead to variation and change in migration-related terminology in EU legal texts.

4.5 Discussion

The findings from this stage of the analysis offer evidence to support the hypothesis that translation-induced language change can occur within the EU language environment. The results show various instances of terminological change, mainly in Italian and to a lesser extent in French, as a result, at least in part, of contact with English in translation. These results therefore corroborate and build upon previous research that has produced similar findings in different contexts (e.g. Malamatidou, 2016; Bisiada, 2016). It is interesting to set the results of this stage of the analysis in the broader context of translation-induced language change research by adopting Johanson's code-copying framework, as outlined in chapter 2. Although more commonly applied to broader, more general language communities and users of everyday language, the 'language community' in this case can be considered to be the legal translators and drafters working on EU legal texts. Indeed, in light of the findings in this chapter, it would appear that for a number of concepts and the terms within them, the process of contact-induced language change may be a significant way along the continuum described by the code-copying framework for this language

community. Indeed, terms in conceptual categories such as LONG-TERM RESIDENT appear to have undergone change attributable in part to language contact through translation and can therefore be deemed to have become 'conventionalised'. This means that linguistic forms such as '*soggiornante di lungo periodo'* in Italian and '*résident de longue durée'* in French have become fully integrated within the particular speech community of the legal drafters and translators in the EU institutions. They have now become the reference terms used by that group to express this specific concept. Another example of this phenomenon can be seen with the concept IRREGULAR MIGRANT. Although not yet at the stage of conventionalization as with the terms for LONG-TERM RESIDENT because competing alternative terms are still present in the later subcorpus, the terms '*irregular migrant'*, '*migrant en situation irrégulière'* [migrant in an irregular situation] and '*migrante irregolare*' [irregular migrant] appear to have now become 'habitualised' within this language community. In other words, the terminological profiles for this concept are clearly undergoing changes without the predominant terms becoming fully integrated and are likely to continue along the continuum towards conventionalization in due course.

A further point to elaborate on from these results, as mentioned in Section 4.3.2 above, is the finding that the French terminology examined is not influenced by English through translation to the same extent as Italian terminology, since six conceptual categories show evidence of influence from English on Italian terminology, while only three show evidence of such influence on French. Indeed, the qualitative results indicate a greater susceptibility to translation-induced language change among Italian terms than French ones. There are two main reasons which can explain this difference. Firstly, on an EU level, French has been a central

procedural language since the organisation's founding and remained the lingua franca within the organisation until relatively recently, until it was surpassed by English (Mattila 2006; Čavoški, 2017). Despite this, French remains very influential within the organisation and continues to hold sway with regards to terminological decisions and primary term creation (Fischer, 2010). It also remains a key working language among civil servants across the EU institutions. Indeed, according to extensive corpus-based research of EU texts, traces of the previously dominant role occupied by French 'are found at the lexical level, especially in English and Italian Eurolects' (Mori, 2018c, pp. 385). Indeed, the continued influence of French can be observed in various lexical terms used in EU English, such as '*stagiaire*', '*domiciliataire*' and '*jurisconsult*', which are obviously of French origin (Kermas, 2010, pp. 55/56). The Italian language does not occupy the same prestigious status as French in the EU institutions, thus making it more susceptible to influence from others.

Secondly, in more general terms, there is particular concern in France about the status of their language in EU politics, given the diminishing status of French both globally and within the EU institutions themselves (Ringe, 2022, pp. 87). Indeed, the French language has particularly strong links to 'a sense of national identity' and 'pride in French culture' (Ban, 2013, pp. 205-6), and has a long history of linguistic prescriptivism. Previous research in terminology has found that 'French tends to create new expressions rather than adopt foreign words, possibly in line with the country's linguistic policy of preservation of the national language' (Bertaccini, Massari & Castagnoli, 2010, pp. 14). This protectionist approach is embodied by the institution of the *Academie Française*, which serves as an authority on matters pertaining to the French language and aims to prevent the adoption of certain

'harmful' or 'avoidable' anglicisms (Academie Française, 1999). In addition, a great deal of French legislation has been passed with the aim of limiting external linguistic influences on the French language, such as the '*Loi Toubon*' (Law no. 94-665 of 4 August 1994). Legislation was also passed in France to set up a commission for the enrichment of the French language (Decree no. 96-602 of 3 July 1996 on the enrichment of the French language), tasked with finding ways of expressing technical and scientific innovations in French rather than borrowing from English.

No such institutions or legislation exist for the Italian language to guard against such influences, and Italy lacks a regulated system for creating and accepting neologisms like in France (Resi & ten Hacken, 2023)[12]. Despite public awareness of the impact of English on Italian, negative reactions to such a phenomenon appear much less widespread than for French (McLaughlin, 2013, pp. 456). Indeed, Italian has been observed to be generally more flexible than French in accommodating and adopting terms from other languages, principally English (Rogato, 2014). This deeply entrenched and long-standing tradition of linguistic protectionism in French culture is also likely to play a role in making French terminology more resistant to certain influences than Italian. This difference is encapsulated by the recent example of the European Green Deal launched in 2019 (*European Council Conclusions*, 12 December 2019). While the French title given to this set of policy initiatives is *Le pacte vert pour l'Europe*, the Italian version given by the European Commission, *Green Deal europeo*, stays much closer to the English title, highlighting the different approaches taken when coining new terms in the two languages on the basis of an English original.

[12] Although a draft bill has recently been proposed in the Italian parliament to set up a 'committee for the protection, promotion and development of the Italian language' (Draft bill C.734, 2022).

While the main objective of this research is to examine the occurrence of translation-induced language change at the EU level and investigate the extent to which this subsequently impacts more general speech communities, it is essential to acknowledge and explore the other factors which are likely to be involved in the language changes that have been identified by this research. Indeed, it is almost always the case that language changes are the result of multiple causes. Aitchison argues that causality must be explored on numerous levels and that 'the immediate trigger must be looked at alongside the underlying propensities of the language concerned' (2013, pp. 202). One of the most important of these factors is the indirect influence, not through the process of translation itself, of English on French and Italian, which could well play a contributing to the changes detected here. This indirect influence takes a number of forms. On a micro level, it has been noted that aside from actual drafting of legal documents, a 'conceptual thinking' also tends to be carried out predominantly in English in the EU environment (Fischer 2009, pp. 73). The fact that general discussions regarding legislation is carried out in English therefore increases its influence on other language versions yet further.

In addition to this, as mentioned in chapter 2, the terminology advocated by instructions for drafters and translators set out in glossaries and databases such as IATE and the EMN glossary is often the subject of primary term creation in English (Fischer, 2010). The result of this is that the terminology used in other languages is inevitably affected by the language from which the original term was created. Although not a direct case of translation-induced language change, the fact that certain EU recommended terms originate in English and are subsequently rendered into other official languages means that English nonetheless has an impact on terminology in other languages in this context. As a consequence, the terminology in

the official EU languages that appears in institutional term bases are often ultimately derived from primary terms created in English. It is important to underline the importance of these terminology databases. Indeed, IATE, the largest terminology database in the world today, covers more than eight million terms across the 24 official EU languages, and in 2020 it received 36.5 million queries in its public version and 21 million in its internal institutional version, highlighting its high level of usage. Moreover, 31% of the entries in IATE also offer 'added value' through the inclusion of context and notes to provide greater detail (Serlet, 2023). There are also other systems in place to ensure consistency of terminology in the EU and minimise terminological instability and variation. The Terminology Coordination unit, known as TermCoord, assists translation units in 'terminology management' and provides 'terminology and documentation tools and resources' (Terminology Coordination Unit, 2018). The European Migration Network also provides its own glossary specifically focusing on migration-related terminology, as described above.

Nevertheless, previous research analysing institutional term bank entries and translation patterns has noted that institutional translators do not always systematically follow terminological recommendations (Prieto Ramos & Morales Moreno, 2019, pp. 108). This may be in part due to the excessive 'noise' in the IATE database (Prieto-Ramos & Guzman, 2023, pp. 251) as well as its acknowledged 'reliability fluctuations' (ibid, pp. 247). The IATE database and available manual dictionaries have been criticised for proving deficient with specific regard to migration terminology (Brannan, forthcoming). Various other persistent problems have also been identified with institutional resources, including excessively large translation memories and inconsistencies between databases (Biel, 2014, pp. 276), which must inevitably cause and compound terminological variation of the kind detected in this

study. Indeed, although these institutional term banks and glossaries undoubtedly play a role in shaping and standardising the terminology used in EU legal texts, this is insufficient in explaining many of the changes detected here, since there are various examples of divergence between the terms prescribed by EU databases and glossaries and terms actually used in the texts.

Another key factor that is important to take into account is the increased use of Computer-Assisted Translation (CAT) tools that have become a central part of translation practice in recent years, including in the EU institutions. Although used principally as a processing tool prior to human translation, the Commission has its own machine translation web service known as eTranslation, and translators in the Parliament also use machine-translated segments, automatically matching new texts with translation memories containing previously translated segments of text (Ringe, 2022, pp. 107). The current trend among multilingual institutions is to integrate term banks, like the ones discussed above, into the organisation's CAT tools (Prieto-Ramos & Guzman, 2023). The fact that previously chunks of text are reused through this process helps to consolidate any changes that occur and thus accelerate the integration of terms into a language, leading to their predominance.

On a more macro level, despite the EU's commitment to cultural and linguistic diversity, the ubiquity of English in all political, social and economic contexts means that it is bound to influence linguistic aspects of French and Italian (and all other official EU languages) on various different levels (Anderman & Rogers, 2005). This influence has been shown to have fundamental effects on a range of linguistic areas in both Italian (e.g. Musacchio, 2005) and French (e.g. Rollason, 2003). These factors of indirect influence, which are not directly attributable to translation, must be taken into account as additional contributing factors to the changes observed in this

analysis. Moreover, another factor that could be attributed to some of the changes identified above is the Principle of Least Effort (Zipf, 1949), which is the hypothesised linguistic tendency to produce shorter-length utterances wherever possible. Indeed, this may be a contributing factor to the increased use of Italian terms '*rimpatriato*' and '*migrante irregolare*', since these terms offer a shorter and more linguistically efficient alternative for expressing the respective contexts. In this regard, some institutional translators have noted that English tends to be particularly concise and accurate, making English-inspired terminological solutions attractive to other languages (Serlet, 2023). The pressure for linguistic efficiency or less effort may well therefore facilitate or work in tandem with the changes induced by contact in translation in these cases.

Also important to note are the different translation processes used by the EU for different legal document types. Binding legislation is subject to multi-stage and multilingual scrutiny and quality assurance that is predominantly carried out in-house, while non-binding texts are often outsourced to external contractors and agencies, and may not therefore be subject to the same strict controls (Biel & Kozbial, 2020, pp. 75-6). This hierarchical system of importance between different text types may well also contribute to some of the discrepancies observed in this study.

In sum, as with all studies of language change, it is unquestionable that there is a complex network of competing influences and causes at play with regard to the changes observed in the corpus analysis carried out at this stage of the study. The evidence collated here allows translation-induced language change to be identified as one causal factor, but others must also be taken into account. The complexity of this network of influences is clear from the fact that there are differences in the way

the languages in question respond to such pressures and discrepancies also in the way in which terms across different conceptual categories behave over time. After all, if the same causal factor or factors were equally responsible for the changes observed here, we might expect uniform changes to occur across the languages and terms being studied here, but this is clearly not the case. Indeed, it appears that the series of factors explored in this discussion combine in different ways with the factor of contact through translation to bring about a range of changes in the different languages and conceptual categories studied. The following two chapters examine how closely these changes are reflected by migration-related terminology used in news articles.

5. Stage 2: *MigrationInTheNews Corpus* analysis

The purpose of this section is to analyse the migration terminology used in English, French and Italian newspaper texts over the same time periods as used for the *EUMigrLaw Corpus* analysis (1992-1998 and 2015-2018). As with the previous stage, this chapter presents the specific methodology used for this part of the study, before presenting the results and analysis, followed by a discussion. As set out previously, this stage seeks to ascertain the degree of observable synchronic and/or diachronic terminological variation within each of the languages in question in migration-related news articles. The aim is to understand how migration-related terms are used differently in each language both concurrently and over time through a combined quantitative and qualitative data analysis. In addition, the patterns of terminological variation observed across the three languages can be analysed to draw conclusions on similarities and differences between the different languages in this context. As explained in chapter 3, the corpus is made up of articles from source newspapers that span the political spectrum with regard to their perspectives on migration. It is to be expected that different terminology will be employed by different newspapers depending on their political stance, but a specific comparison between individual news outlets falls outside the scope of this study.

5.1 Methodology

Following the compiling of the *MigrationInTheNews* C*orpus* as described in chapter 3, all migration-related terms for people need to be extracted from the corpus in exactly the same way as was done for the *EUMigrLaw Corpus* as per chapter 4 (4.1). Through this method, single-word and multi-word terms for people involved in

migration are obtained from the newspaper corpus for each of the three languages. They are then placed into conceptual categories as per the methodology set out in chapter 4 (4.1), accompanied by the absolute frequencies and relative frequencies (per million words) of each individual term. This data can subsequently be used to conduct calculations to produce information about each term and conceptual category.

Firstly, on a conceptual level, the prevalence of each conceptual category as a proportion of the overall number of terms can be calculated by dividing the number of terms recorded in each category by the total number of terms recorded for that time period. This produces a percentage showing how prevalent each concept is in the corpus for both time periods. These percentages indicate which concepts appear with the highest frequency across the corpus and those concepts which represent over 1% as a proportion of the overall number of terms across the three languages are subsequently selected for a more in-depth analysis.

This analysis involves an exploration of the terms contained within each conceptual category for each time period and language. Since one of the aims of this stage is to examine diachronic terminological variation, the relative frequency of each term in the earlier and later subcorpora can be used to calculate the percentage change of that term over time. This can be done using the following calculation:

Percentage change = (difference between frequencies/frequency in earlier subcorpus) x 100.

Evidently, this can produce a positive or negative percentage depending on whether the relative frequency of a term has increased or decreased over time. These percentages therefore provide comparable values that indicate both the magnitude

and the direction of the change in frequency of each term across the time period. Those cases where a term only appears in one of the subcorpora and not the other can be considered special cases. For such terms, the percentage change value is not given, since the introduction of a 0 into this calculation produces extreme values which do not allow for a representative comparison. These cases are examined on an individual basis and not represented in the graphs below.

The percentage change values calculated in this way can subsequently be subjected to statistical significance testing. In order to do this, the log likelihood value can be used. The log likelihood value is commonly used in corpus linguistics to determine the level of statistical significance of the difference in relative frequency of a given term across two corpora (Rayson & Garside, 2000). This value can be calculated using an online log-likelihood wizard (Rayson, 2016). This platform allows the researcher to enter the absolute frequency of a term in two different (sub)corpora along with the size of each (sub)corpora in order to calculate the log likelihood value for a given term. The Fisher's exact test and SPSS cross-tabulation detailed in chapter 4 is not necessary for this part as the values investigated here are much larger[13]. As is customary in corpus linguistic studies, the cut-off value for statistical significance used in this study is $p<0.05$ (e.g. Brezina, 2018, pp. 12). The p-value of $p<0.05$ equates to a log likelihood value of 3.84. Therefore, all cases where the log likelihood value is higher than 3.84 can be deemed statistically significant. It is useful to combine both these complementary metrics of percentage change and log likelihood, since a change can be statistically significant, but only very small in size.

[13] As described in chapter 4, the Fisher's exact test is used when expected frequencies in a cross-tabulation are particularly small, which is not the case here.

In addition, the prevalence of each individual term within each conceptual category can be calculated as a proportion of the total number of terms in that category in order to provide more information on the internal dynamics within each category. This is a valuable metric for ensuring a comprehensive picture of the terminological changes that have occurred, since the frequency of a term in a given category may decrease overall, but simultaneously increase as a proportion of that category. Table 16 below provides an example of the data that can be collated following the above methodology. The table shows the relative frequency of each term in the conceptual category THIRD-COUNTRY NATIONAL for Italian, the percentage change in frequency of each across the time period and the proportion of each term in the earlier and later subcorpora. Also, based on the statistical significance test described above, those percentage changes deemed statistically significant are highlighted in bold. All the results tables for each concept representing at least 1% of the overall number of terms are provided in appendix 2.

Table 16: THIRD-COUNTRY NATIONAL conceptual category: Italian

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Straniero [Foreigner]	726.64	659.44	-9.25%	50%	88%
Cittadino straniero [Foreign citizen]	28.97	41.22	42.29%	2%	6%
Extracomunitario [non-EU]	663.87	30.91	**-95.34%**	45%	4%
Cittadino extracomunitario [Non-EU citizen]	36.21	15.46	-57.30%	2%	2%
Non europeo [Non-European]	7.24	0	**N/A**	>1%	0%

5.2 Quantitative results

This chapter firstly examines the general quantitative observations from the newspaper corpus analysis, and then investigates in more detail the terminological trends in a range of conceptual categories.

Table 17: Number of terms in MigrationInTheNews Corpus by language and time period

	Number of terms in *MigrationInTheNews Corpus* 1992-1998	**Number of terms in *MigrationInTheNews Corpus* 2015-2018**
English	36 Mean: 2.77 terms per concept Median: 2 Mode: 1	51 Mean: 3.33 terms per concept Median: 2 Mode: 1
French	43 Mean: 2.93 terms per concept Median: 1 Mode: 1	55 Mean: 3.13 terms per concept Median: 1 Mode: 1
Italian	36 Mean: 2.92 terms per concept Median: 2 Mode: 1	39 Mean: 2.92 terms per concept Median: 2 Mode: 1

As shown in table 17 above, the results for English show a total of 36 terms in the 1992-1998 time period, which can be classified into 13 conceptual categories. The mean number of terms per concept for this period is therefore 2.77, while the median number of terms per concept is 2 and the mode is 1. This indicates that the relatively high number of overall terms to concepts is due to a small number of categories containing a high number of terms. Indeed, the conceptual categories ASYLUM SEEKER, IRREGULAR MIGRANT, MIGRANT and THIRD-COUNTRY NATIONAL all contain 5 terms each in this period. Meanwhile, five of the conceptual categories only contain a single term. There is therefore a very wide variety in the number of term variants per concept depending on the specific category in question in this earlier period. The terminological variation in migration-related terminology becomes

even greater in news articles in the later period, with the mean number of terms per concept increasing to 3.33 terms per concept. These results for English imply that rather than converging towards an accepted set of terms for migration concepts, the number of term variants to express different concepts is in fact growing.

Indeed, the results show that there has been a very significant change in the use of migration terminology in English newspapers over the time span being studied. Comparing all the terms used across both periods, the log likelihood values indicate a statistically significant difference in frequency in 52 cases, while only 7 terms show a difference in frequency which cannot be deemed statistically significant. This means that 88% of the English terms identified across the corpus show a statistically significant difference in frequency between the two time periods. Among those terms which experienced a significant change in frequency, 31 of them increased in frequency between the time periods, while 21 decreased in frequency.

The results from the 1992-1998 subsection of the French part of the corpus show a total of 43 terms, which can be classified into 14 categories. The mean number of terms per concept is therefore 2.93, which is slightly higher than the mean number of terms per concept in the English part of the corpus for this period. However, the median number of terms per concept is 1, which is lower than for English. This indicates that a few extreme outliers are driving up the mean, in particular IRREGULAR MIGRANT (14 terms). Meanwhile, four of the conceptual categories in this period only contain a single term, underlining the significant spread in the number of terms depending on the concept in question. Like English, French also shows a rise in the mean number of terms per concept over the time period, growing to 3.13. However, this increase is not as extreme as it is for English and the median

and mode values both remain at 1, suggesting that the number of term variants has increased, but not to the same degree as for English.

The French results indicate a significant change in the use of migration terminology in French newspapers over the time period. The log likelihood values show that there is a statistically significant difference in the frequency of terms in 48 cases, while 19 cases show no statistically significant change. This equates to 72% of terms overall showing a statistically significant difference in frequency between the earlier and later time periods. Among those terms which underwent a statistically significant change in frequency, 30 terms increased in frequency, while 18 decreased in frequency. Interestingly, the percentage of terms which show a significant change is lower than for the English part of the newspaper corpus, which again suggests that the change undergone by French migration terminology is not as extreme as experienced by English migration terminology. Indeed, this result would suggest greater levels of stability in terminology used by French newspapers in relation to migration topics than English newspapers.

Italian is the only language which does not demonstrate an increase in the mean number of terms per concept during the time period. Indeed, this value remains stable at 2.92 for both periods, with the median and mode values also remaining at 2 and 1 respectively. Once again, these results indicate a small number of categories with a high number of terms, which are pushing the mean value up. Indeed, there is only one category with a high number of terms, IRREGULAR MIGRANT, containing 9 term variants in both periods.

In common with the other languages, the Italian results show a significant change in the migration terminology used in Italian newspapers. According to the log likelihood

values, there is a statistically significant difference in the frequency of terms in 43 cases, while 12 cases show no statistically significant change. This equates to 78% of terms overall showing a statistically significant difference in frequency over the time period. Among those terms which experienced a significant change in frequency in Italian, 22 showed a significant increase while 21 recorded a significant decrease in frequency.

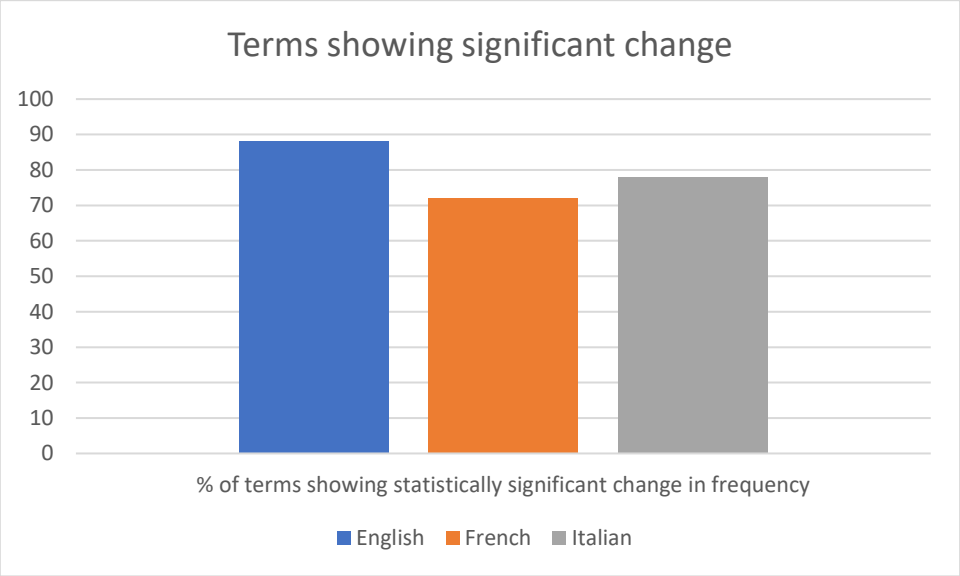


Figure 9: percentage of terms showing a statistically significant change in frequency

Taken together, based on overall percentages of statistically significant change, English is therefore the language which experienced the most significant degree of terminological change between the two time periods, followed by Italian, then French, as shown by figure 9 above.

For all three languages, there are also more terms which underwent a significant increase than those that experienced a significant decrease. This uniform result across the three languages is all the more interesting given the significant differences between the languages with regard to the changes in the overall number of terms identified in both time periods. Indeed, the overall number of terms identified

across all categories in the English part fell from 2,968 to 2,681 (7,787.66pmw to 7,033.79pmw). This equates to a percentage change of -9.68%, which can be deemed statistically significant (log likelihood 11.58). This overall reduction is largely a result of significant falls in the number of terms in the conceptual categories ASYLUM SEEKER (351 fewer terms), REFUGEE (272 fewer terms) and IRREGULAR MIGRANT (180 fewer terms).

In contrast, in the French part of the corpus, the relative frequency of terms identified increased to a statistically significant extent from 2,243 to 2,682 (5,668.99pmw to 6,650.09pmw) (% change: 17.31%, log likelihood: 39.77). This is largely due to a very significant increase in the number of terms in the conceptual category MIGRANT (509 more terms). In the Italian part, the overall number of terms is very consistent, with 2,311 in the earlier subcorpus compared to 2,295 in the later one (5,578.92pmw to 5,911.79pmw), equating to a percentage change of 5.97%, which cannot be deemed a statistically significant change (log likelihood: 0.04). As a result of these changes, the French part of the corpus changes from being the language with the fewest overall terms in the earlier time period to the one with the most in the later one.

Table 18 below shows each conceptual category in each language as an overall proportion of all migration terms identified in both time periods.

Table 18: Overall proportions of concepts

Category	English		French		Italian	
	Proportion in Earlier Subcorpus	Proportion in Later Subcorpus	Proportion in Earlier Subcorpus	Proportion in Later Subcorpus	Proportion in Earlier Subcorpus	Proportion in Later Subcorpus
ASYLUM SEEKER	**14.56%**	**3.02%**	**4.32%**	**5.07%**	0.09%	**6.01%**
DETAINEE	0.61%	0.26%	0.13%	0%	0%	0%

DISPLACED PERSON	0%	0%	0.09%	0.15%	**2.16%**	0.31%
INTERNATIONAL STUDENT	0.51%	**2.98%**	0.40%	0.41%	0%	0.13%
IRREGULAR MIGRANT	**16.04%**	**10.67%**	**23.27%**	**10.89%**	**20.68%**	**7.49%**
MIGRANT	**39.18%**	**56.40%**	**26.35%**	**41.01%**	**39.51%**	**49.06%**
MIGRANT WOMAN	0%	0%	0%	0.11%	0.13%	0.13%
MIGRANT WORKER	**1.48%**	**4.20%**	**2.18%**	**1.60%**	**1.08%**	**1.66%**
NON-DEPORTABLE PERSON	0%	0%	0.13%	0%	0%	0%
OVERSTAYER	0.84%	0.34%	0%	0%	0%	0%
POSTED WORKER	0%	0%	0%	0.19%	0%	0%
REFUGEE	**16.91%**	**8.58%**	**11.28%**	**14.65%**	**8.35%**	**19.43%**
REJECTED APPLICANT FOR INTERNATIONAL PROTECTION	0.17%	0.48%	0.53%	**1.75%**	0%	0%
RETURNEE	0.47%	0%	0%	0%	0.22%	0%
SEASONAL WORKER	0%	0.15%	0.13%	0.11%	0.36%	0.04%
STATELESS PERSON	0%	0%	0%	0.26%	0%	0%
THIRD-COUNTRY NATIONAL	**8.12%**	**3.66%**	**30.00%**	**19.16%**	**26.22%**	**12.64%**
UNACCOMPANIED MINOR	0%	0.41%	0.04%	0.19%	0%	0.26%
UNION CITIZEN	**1.18%**	**7.68%**	0.89%	**4.81%**	**1.25%**	**2.66%**
VICTIM OF TRAFFICKING IN HUMAN BEINGS	0%	0%	0%	0%	0.04%	0%
VULNERABLE PERSON	0%	0.48%	0%	0.04%	0%	0%

Having examined the general quantitative trends which emerge from the corpus analysis, the sections below examine the intralingual and interlingual trends across a

selection of the conceptual categories under investigation. Those concepts which represent a proportion of at least 1% or more of the migration terms in all three languages in the earlier and/or later corpus are highlighted in bold in the table above and are selected for more in-depth analysis below. The below analysis in 5.3 below includes graphs to show the percentage changes in the frequency of each term which has increased or decreased to a statistically significant extent. It is also important to reiterate that percentage change values where a term does not appear at all in one of the time periods (i.e. has a frequency of 0) are unrepresentative, therefore those terms are not included in the graphs and percentage change values are not provided for those terms. Those terms are nonetheless taken into account in this analysis and are discussed in each case.

5.3 Terminological trends in conceptual categories

5.3.1 ASYLUM SEEKER and REFUGEE

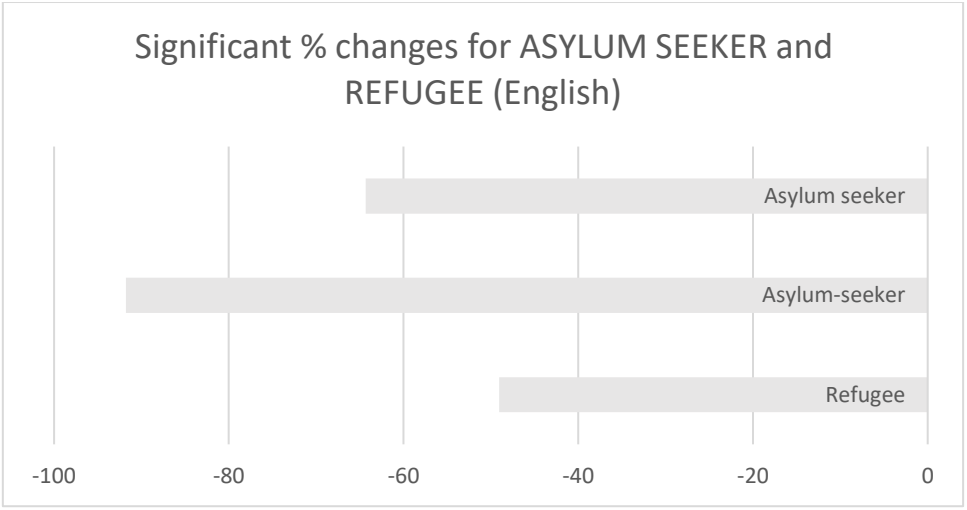


Figure 10: Percentage change in relative frequency of terms in conceptual categories ASYLUM SEEKER and REFUGEE for English news articles

It is interesting to observe that, for English, two central concepts in the field of migration, ASYLUM SEEKER and REFUGEE, both fall dramatically as a proportion of the overall number of terms over the time period, with ASYLUM SEEKER falling from 14.56% to 3.02% and REFUGEE dropping from 16.91% to 8.58%, as shown in table 18 above. In the category ASYLUM SEEKER, all terms see a decrease in relative frequency, with three of the five terms, '*people seeking asylum*', '*applicant for asylum*' and '*asylum applicant*', falling out of the corpus completely. The remaining terms, '*asylum seeker*' and '*asylum-seeker*'[14], record very substantial percentage decreases at -64.32% and -91.78% respectively, as shown in figure 10 above. Similarly, as the only term in the category REFUGEE, the term '*refugee*' experiences a very significant drop in relative frequency (-49.05%). The decrease in these categories as a proportion of the overall terms indicates that rather than older terms being replaced by newer term variants to express the same concepts, these concepts are simply less frequently referred to in the later subcorpus. This finding is corroborated by other research, which shows that the UK press has an increasing preference for the use of the more general term 'migrant' as opposed to the more humanitarian-focused terms of '*asylum seeker*' and '*refugee*' (Berry, Garcia-Blanco & Moore, 2015). This is particularly striking given that the later subcorpus covers a time period which saw extremely high numbers of refugees entering Europe, fleeing war and persecution, mainly from Syria, Afghanistan and Iraq (European Commission, 2017). This therefore demonstrates that even seemingly major events do not always dictate the extent to which certain terminology is used. The fall in the use of the terms '*asylum seeker*' and '*refugee*' appears to result from a trend in some UK

[14] Orthographic variants are considered as different terms for the purpose of this study, as per Faber & León-Araúz (2016, pp. 12-13).

newspapers to focus on those supposedly less deserving of aid, while shifting the focus away from other people in a more desperate situation, or to place these two groups in opposition to each other, as illustrated by the following examples:

> *How can we control migrant numbers but help real **refugees**?*
>
> (Evening Standard, 2015)
>
> *Many migrants flooding into the UK are doing so for economic reasons and not for asylum.*
>
> (The Express, 2015)

This approach is used largely in tabloid or mid-market newspapers to reduce empathy towards those arriving from other countries and, in extreme cases, to demonise them. Moreover, the frequent collocation of negative or catastrophic imagery, as per the use of 'flooding' in the above example, adds to the sense of moral panic connected to those migrating into the UK (Taylor, 2014).

It is also the case that the print media coverage during the later time period coincides with the Brexit referendum, which focused heavily on migration from Europe, rather than asylum seekers from outside the EU. Therefore, the significant decrease in the prevalence of the concepts ASYLUM SEEKER and REFUGEE is not a consequence of alternative terms being used to express the same concepts but instead results from a shift in focus by many of the newspapers included in the corpus, often motivated by their objective to generate a negative narrative around the topic of migration.

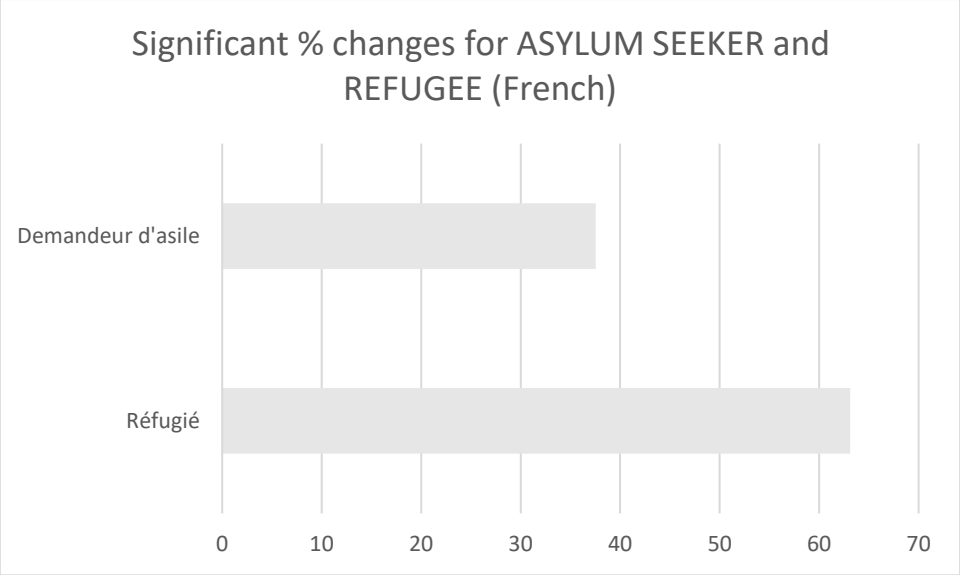


Figure 11: Percentage change in relative frequency of terms in conceptual categories ASYLUM SEEKER and REFUGEE for French news articles

In contrast to English, both conceptual categories ASYLUM SEEKER and REFUGEE in French increase as a proportion of the overall number of terms. Indeed, the category ASYLUM SEEKER increases from 4.32% of the overall number of terms in the earlier subcorpus to 5.07% in the later one, while REFUGEE increases from 11.28% to 14.65%, as shown in table 18. These two categories also demonstrate a level of terminological stability which is rare in this corpus study with both categories only containing a single term in both time periods. Indeed, although English also has only one term variant in both subcorpora for the category REFUGEE, it has five term variants for ASYLUM SEEKER, while Italian has two term variants for ASYLUM SEEKER and three for REFUGEE. The statistically significant increase in relative frequency of both terms in French suggests greater coverage in the French press of these topics over time. This is unsurprising given the number of asylum seekers entering Europe during the later time period and reflects the heightened reporting on the plight of those particular migrants.

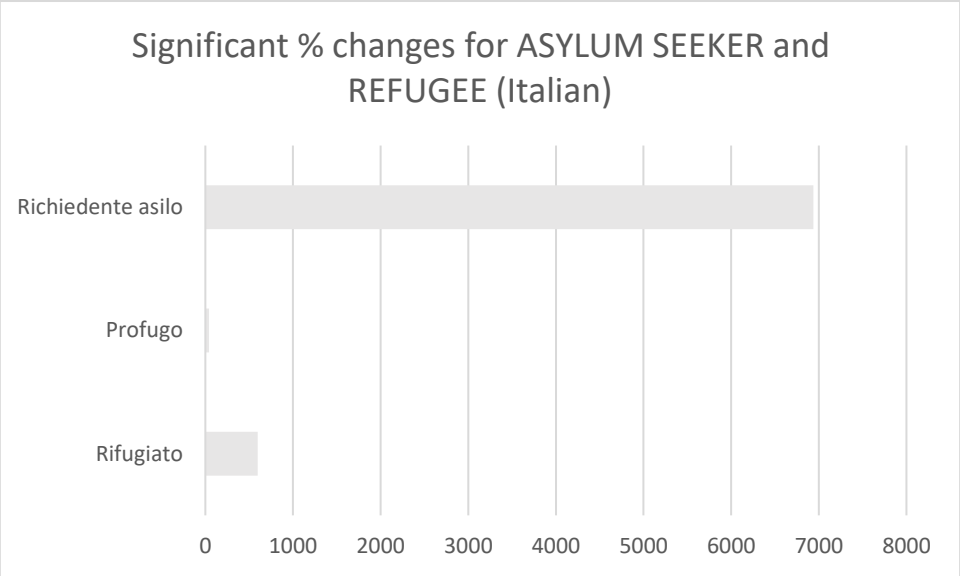


Figure 12: Percentage change in relative frequency of terms in conceptual categories ASYLUM SEEKER and REFUGEE for Italian news articles

For Italian, the proportional increase of the categories ASYLUM SEEKER and REFUGEE is even more substantial than for French, as shown in table 18. The contribution of the concept ASYLUM SEEKER in the Italian part in the earlier subcorpus is negligible, equating to 0.09%, but this jumps to 6.01% for the later subcorpus. Likewise, the concept REFUGEE accounts for 8.35% of the total number of terms in the earlier subcorpus and increases to 19.43% in the later subcorpus. All statistically significant relative frequency changes are substantial increases, as shown in figure 12, with '*richiedente asilo*' [asylum seeker] (6,939.75%) and '*rifugiato*' [refugee] (597.92%) constituting two of the highest percentage increases across all Italian migration terms in the corpus.

Moreover, the fact that both terms, '*rifugiato*' and '*profugo*', show a significant increase in frequency suggests that the specific topic of refugees within the broader

theme of migration received greater coverage in the recent time period. Rather than a conscious focus on this topic by the newspapers, this is likely to be the result of the fact that the later time period coincides with the highest migration flow into Europe since World War II, with around 150,000 migrants and asylum seekers arriving in Italy in 2015 alone and almost 3,000 drowning along the perilous Mediterranean migration route from North Africa (IOM, 2015). Indeed, the International Rescue Committee estimates that over 700,000 migrants and asylum seekers arrived in Italy between 2014-2020, a period partially covered by the later subcorpus, and points out that Italy receives more asylum seekers per year than any other EU country (IRC, 2022). The topic of refugees and asylum seekers was therefore particularly relevant in Italy during this time period, given the specific 'point of arrival' role it plays within the EU.

Another interesting trend to examine in the Italian context is the change in the usage of the terms '*profugo*' and '*rifugiato*'. In Italian newspapers, these terms are often used interchangeably to express the concept of REFUGEE. However, as mentioned in the previous chapter, only the term '*rifugiato*' expresses the strict legal definition of refugee as enshrined in the Convention Relating to the Status of Refugees, also known as the Geneva Convention, meaning 'someone who is unable or unwilling to return to their country of origin owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion' (Convention relating to the Status of Refugees, 1951). The term '*profugo*' is a more general term referring to a person who has fled their home country for any reason whatsoever, yet does not necessarily fulfil the criteria to obtain refugee status. In the corpus analysis, both terms underwent a statistically significant increase in frequency over the time period, although the term '*rifugiato*'

increased in frequency much more dramatically than '*profugo*', as indicated by their corresponding percentage change values shown in figure 12 above. Indeed, '*rifugiato*' recorded a percentage change value of 597.92% compared to 42.27% for '*profugo*'. Therefore, by the time period covered by the later subcorpus, '***rifugiato'*** has become the most frequently used term for the concept REFUGEE, overtaking '*profugo*', and constituting 54% of the overall category by the later period. This trend indicates a growing awareness of Italian newspapers regarding the importance of terminological accuracy, highlighted by the fact that the terms '*rifugiato*' and '*profugo*' often appear alongside one another, as in the following example taken from the later subcorpus:

> …*associazioni e altre realtà che a vario titolo si occupano di stranieri,* ***profughi*** *e* ***rifugiati***
>
> […*associations and other organisations which deal variously with foreigners,* ***profughi*** *and* ***refugees***]
>
> (Il Giorno, 2018).

Such usage shows an acknowledgement of the semantic distinction between the two terms and an awareness that they should not be used interchangeably.

In sum, there is a very striking difference in the frequency trends for the concepts REFUGEE and ASYLUM SEEKER between English, on one hand, and French and Italian on the other. While English sees a very significant decrease in the overall prevalence of both these concepts over the time period, French and Italian record substantial increases. This indicates a significant difference in the focus of the print media in these countries with regard to the topic of migration. Despite this difference, all three languages display relative terminological stability for these concepts with

predominant terms remaining the most prevalent or becoming even more so, with the exception of the concept REFUGEE in Italian. In this specific case, the more legally appropriate term '*rifugiato*' overtakes the general term '*profugo*' to become the most frequent term in that category by the later time period.

5.3.2 MIGRANT

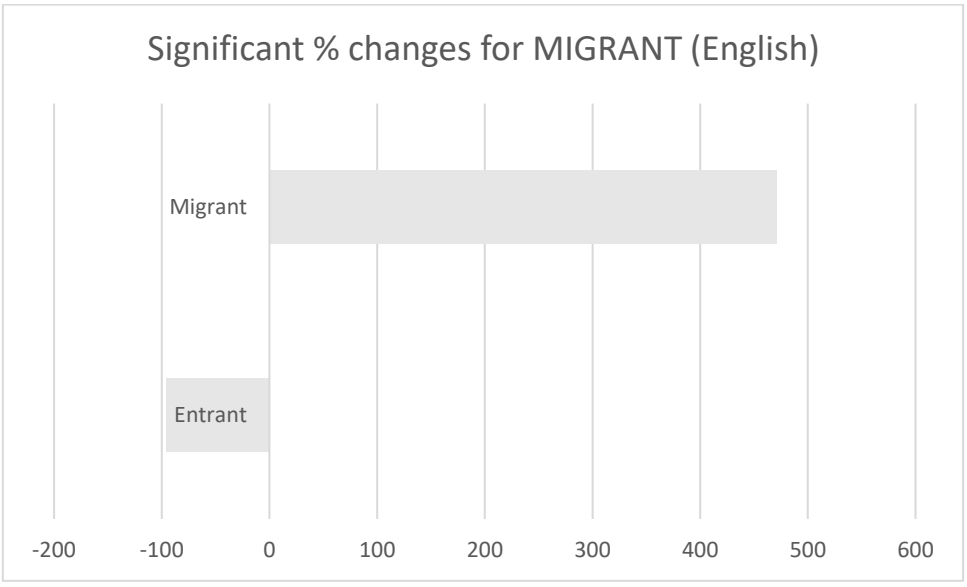


Figure 13: Percentage change in relative frequency of terms in conceptual category MIGRANT for English news articles

In comparison to the overall number of terms identified across all categories in each subcorpus, the conceptual category MIGRANT saw a proportional increase in the English part of the newspaper corpus, containing 39.18% of the total terms in the earlier subcorpus and 56.40% of the total terms in the later subcorpus. The results also show a significant change in the terminology used in this conceptual in the English part of the corpus between the two time periods. As shown in figure 13 above, two out of the five terms in this category recorded a statistically significant difference in frequency between the two subcorpora, namely '*migrant*' which saw a

significant increase and '*entrant*' which recorded a substantial decrease. With its relative frequency increasing from 325.36pmw to 1,860.11pmw, the term '*migrant*' sees a percentage change of 471.71%. Within the conceptual category, it also saw a very significant proportional increase (from 11% to 47%). The term '*entrant*', however, experienced a very significant decrease (-96.45%), with the term almost falling out of usage by the later time period. At the same time, within the same conceptual category, the term '*immigrant*' has seen no statistically significant change in frequency across the time period, with a relative frequency in the earlier time period of 2,235.54pmw compared to 2,049.01pmw in the later one. However, its proportion within the category fell from 73% to 52%. Therefore, in the earlier time period, English newspapers had a clear preference for the use of the term '*immigrant*' to express the concept MIGRANT but the dramatic increase in the occurrence of '*migrant*' by the later time period means that it is used to a very similar extent by the time of the later subcorpus, making up 47% of occurrences in this category, as per the below example:

> ***Migrants*** *have not only filled gaps for skilled roles, they brought with them new ideas of entrepreneurship*.
>
> (Evening Standard, 2015)

The trend in English is therefore towards a significant increase in the use of the term '*migrant*', leading to a proportional fall in the use of '*immigrant*'. Nevertheless, the term '*immigrant'* is still narrowly the most frequently used term in this category, making up 52% of occurrences, as per the following example:

> *In addition, the number of* ***immigrants*** *coming here is offset to a large extent by the number of British citizens living abroad*.

(Times, 2016)

There is no universally recognised legal definition of the terms '*migrant*' and '*immigrant*' and a wide range of definitions are used by different institutions and organisations. The International Organization for Migration (IOM, 2022) defines a migrant as 'a person who moves away from his or her place of usual residence, whether within a country or across an international border, temporarily or permanently, and for a variety of reasons', while it classifies an immigrant 'from the perspective of the country of arrival, [as] a person who moves into a country other than that of his or her nationality or usual residence'. The central semantic difference is that the term '*immigrant*' sees the person in question from the perspective of the country of arrival, while '*migrant*' can refer to a person entering or leaving a country, or moving within a country. The results above therefore suggest a rise in terminological sensitivity from some newspapers demonstrated by the increased use of the semantically broader term '*migrant*'.

The term '*migrant*' also has more neutral connotations in comparison to the negative semantic prosody of '*immigrant*'. Semantic prosody refers to the situation where a given word co-occurs regularly with certain other words, and then acquires some of the meaning of those words as a result (Stewart, 2010, pp. 1). In this case, the frequent co-occurrence in the media of the word '*immigrant*' with negative words (e.g. 'illegal', 'terrorist', 'flood', etc.) over a long period of time has led it to acquire negative connotations. Indeed, according to a report conducted by Allen and Blinder (2013), 'the most common descriptor for the word IMMIGRANTS across all newspaper types is ILLEGAL'. The frequency of this negative collocation along with the widespread use of the term in negative newspaper articles has led to the term '*immigrant*' acquiring negative semantic prosody. The increased use of '*migrant*'

therefore indicates a degree of awareness of the need to use a more neutral term for this general concept.

However, it is also important to acknowledge that while the term '*migrant*' has increased in English news articles, both in absolute terms and proportionately within its conceptual category, and there has been a proportionate decrease in the use of *'immigrant'*, it remains the most frequently used term in this category. There has also been no statistically significant decrease in its frequency in absolute terms in UK newspapers according to this corpus analysis. This indicates that there has only been a partial recognition of the semantic and connotational issues attached to the term '*immigrant*' among the print media. Equally, newspapers with an anti-migration stance may also seek to harness the negative semantic prosody of certain migration terms, such as '*immigrant*', in order to accentuate the negative depiction that they seek to perpetuate around the theme of migration. Moreover, although the increase in the frequency of the term '*migrant*' may suggest greater terminological sensitivity, there are examples in some contexts, especially tabloid and mid-market papers, it has also been noted that media coverage often uses it pejoratively to refer to 'economic migrants' who are 'less worthy' than other categories such as refugees or asylum seekers (Allen & Blinder, 2013, pp. 27). Such usage is explored further in section 5.3.4 below. The increased frequency of the term '*migrant*' cannot therefore be interpreted as a purely progressive terminological trend since the reasons for this increase are rooted in very different motivations.

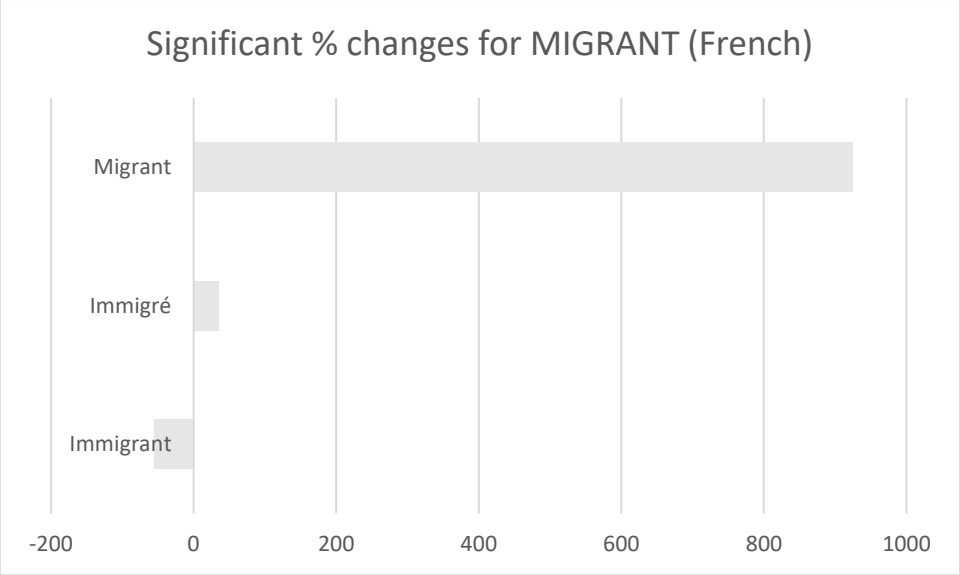


Figure 14: Percentage change in relative frequency of terms in conceptual category MIGRANT for French news articles

As with the English part of the newspaper corpus, the French conceptual category MIGRANT also increased its proportional representation among the overall number of terms identified in the corpus across all categories. The category contains 26.35% of the total number of terms in the earlier subcorpus, which increases to 41.01% in the later subcorpus. Of those French terms in this category which show a significant change in frequency, the term which experienced the most significant increase in frequency was the term '*migrant*', as shown in figure 14 above, with a relative frequency of 141.54pmw in the earlier time period compared to 1450.52pmw in the later one, equating to a percentage change 924.81%. Within this conceptual category, the term '*migrant*' represented only 9% of cases in the earlier time period, which rises dramatically to 53% of cases by the later time period, thus becoming the predominant term for this concept. Meanwhile, in the same conceptual category, the term '*immigrant*' also saw a significant decrease in frequency over the same time period, falling from 588.89pmw to 260.35pmw, with a percentage change of -55.79%, decreasing proportionately in the category from 39% of cases to just 10%. It

is important however to point out that French also has an additional alternative term, '*immigré*', which saw a significant increase in frequency over the time period of 35.70%. However, in proportionate terms within the category, it fell from being used in 46% of overall cases to 34% over the time period. The French terms '*immigrant*' and '*immigré*' both refer specifically to migrants from the perspective of the country of arrival, but they allow for a more nuanced description than is possible with the single English term '*immigrant*'. Indeed, while '*immigrant*', describes a person who is in the process of migrating, '*immigré*' refers to a person who has completed their migration and is now in a settled situation. The below example featuring both these terms is taken from the corpus:

> "*Beaucoup d'**immigrants** viendraient en fait dans l'Hexagone contracter un marriage avec une fille de la deuxième generation d'**immigrés** qui est donc de nationalité française*."
>
> [*Indeed, it appears that many **immigrants** come to France to marry girls who are second generation **immigrants**, and therefore have French nationality*.]
>
> (Les Echos, 1993)

It is therefore interesting to note the significant change in usage between the two terms, with the frequency of the term '*immigrant*' falling significantly, while the term '*immigré*' saw an increase in overall relative frequency in the corpus but a fall in proportional frequency within the conceptual category.

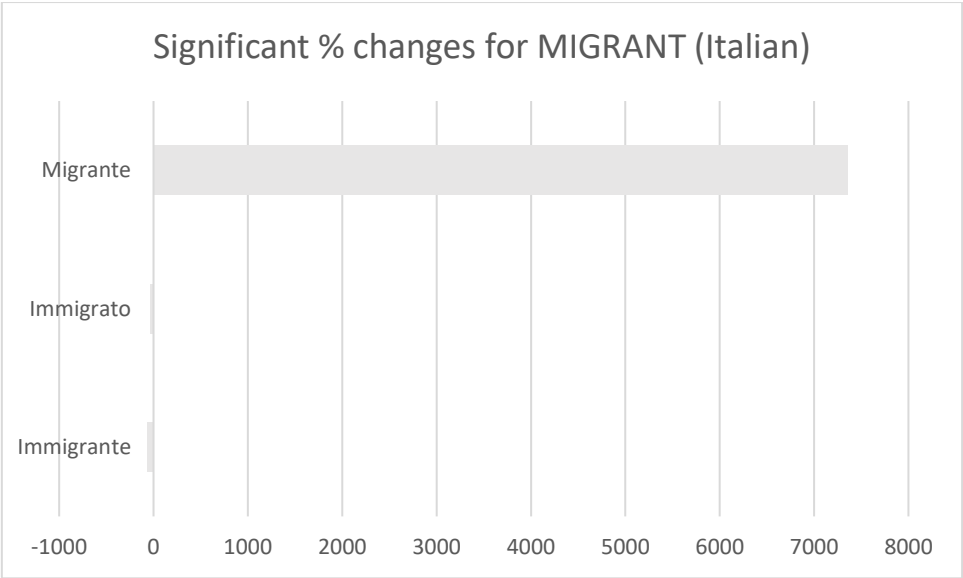


Figure 15: Percentage change in relative frequency of terms in conceptual category MIGRANT for Italian news articles

As observed in both the English and French parts of the newspaper corpus, the conceptual category MIGRANT also increased the proportion of overall terms it contains over the time period. In the earlier subcorpus, it represents 39.51% of the overall terms, compared to 49.06% in the later time period. This is therefore a uniform pattern across the three languages, with the conceptual category MIGRANT increasing in prevalence over the time period with regard to the overall number of terms identified in each language. Also, in common with both English and French terminological trends, the results of the Italian newspaper corpus analysis show significant changes in the terms used in the conceptual category MIGRANT. The term '*migrante*' [migrant] is the only Italian term in this conceptual category which sees a significant increase in frequency, with a huge percentage change of 7,356.37%. With a relative frequency of just 21.73pmw in the earlier subcorpus, this frequency rises very dramatically to 1,620.27pmw. Within the category, this equates to a proportional increase from 1% of the overall terms to 56%. At the same time, the other two term variants in Italian, '*immigrante*' and '*immigrato*' [immigrant], both see

significant decreases in frequency across the time period, with percentage changes of -70.89% and -38.33% respectively. As with French, Italian also has two terms to express the English term ‘immigrant’, where ‘*immigrante*’ describes a person in process of migrating, while ‘*immigrato*’ refers to a person who is more settled in the country of arrival. The term ‘*immigrato*’ sees a proportional decrease within the category from 91% to 43%. Meanwhile, the much less prevalent term ‘*immigrante*’ falls from 2% of cases in the category to just 1%. Once again, it appears that there is a terminological trend in Italian newspapers to shift the perspective away from the country of arrival embodied by the terms ‘*immigrante/immigrato*’ towards a more general label, ‘*migrante*’, referring to people migrating in any direction. Similar to English, the term ‘*migrante*’ also carries a more neutral semantic prosody compared to ‘*immigrato*’ and ‘*immigrante*’, which often appear in collocation with negative adjectives, such as ‘*clandestino*’ or ‘*illegale*’, as per the following examples taken from the corpus:

> “*Occorre poi che siano ben chiari i rischi legati all’accoglienza di* ***immigrati*** *illegali*”
>
> [*It is important to be clear about the risks connected with welcoming illegal* ***immigrants****…*]
>
> (La Nazione, 2018)

> “*se i sindaci intendono riempire le loro città di* ***immigrati*** *clandestini saranno gli elettori a decidere se questi sindaci hanno lavorato bene o male*”
>
> [*if mayors plan on filling their cities with clandestine* ***immigrants****, their electors will decide whether they have done a good or bad job*.]

(Corriere della Sera, 2018).

In this category, the terms referring to a migrant from the perspective of the home country (EN: '*emigrant*', FR: '*émigrant*', IT: '*emigrato*' and '*emigrante*') have undergone different trends in the languages being studied. In English, the term '*emigrant*' only appears in very small numbers in both time periods and the change in frequency is not statistically significant. In French, the term '*émigrant*' is completely absent from the later subcorpus, having represented 2% of the category in the earlier period. However, the most significant change in proportional terms is in Italian, with the terms '*emigrante*' and '*emigrato*' taken together representing 4% of the total terms in the category MIGRANT in the earlier subcorpus. By the later corpus, this frequency has fallen to just 1%, with the term '*emigrante*' completely absent in the later subcorpus. This may be explained in large part by the different historical contexts in the countries in question. Indeed, Italians have accounted for the largest number of voluntary emigrants in recorded history, with 13 million people leaving between 1880 and 1915, and numbers of emigrants remaining very high well into the 1960s (Scotto, 2017). This historical backdrop of large-scale emigration, especially when compared to the UK or France, may explain the higher use of the terms '*emigrato*' and '*emigrante*' in the earlier subcorpus, while the significant drop is likely due to the change in focus from people leaving the country to people arriving from elsewhere as a result of changes in migratory flows.

The results therefore show that in the conceptual category MIGRANT, the terms '*migrant* [EN] / *migrant* [FR] / *migrante* [IT]' have seen a very substantial increase in all three languages. The terms '*migrant*' and '*migrante*' have become predominant in

this category in both French and Italian (53% and 56%). In English, although '*immigrant*' remains the predominant term, there is a clear trend towards the use of '*migrant*', which saw a very substantial proportional increase within the category. At the same time, many of the other terms in this category across the three languages have seen significant decreases in usage, including '*immigrato*', '*immigrante*' and '*emigrato*' in Italian, '*immigrant*' and '*émigrant*' in French and '*entrant*' in English. Even those other term variants which did not register a significant fall in their overall relative frequency in the corpus, such as '*immigré*' in French and '*immigrant*' in English, saw a substantial decrease in their proportional frequency within the category. These results therefore show a uniform shift from newspapers in all three languages towards the general term '*migrant*' and away from alternative term variants in the category which had been more prevalent previously.

5.3.3 IRREGULAR MIGRANT

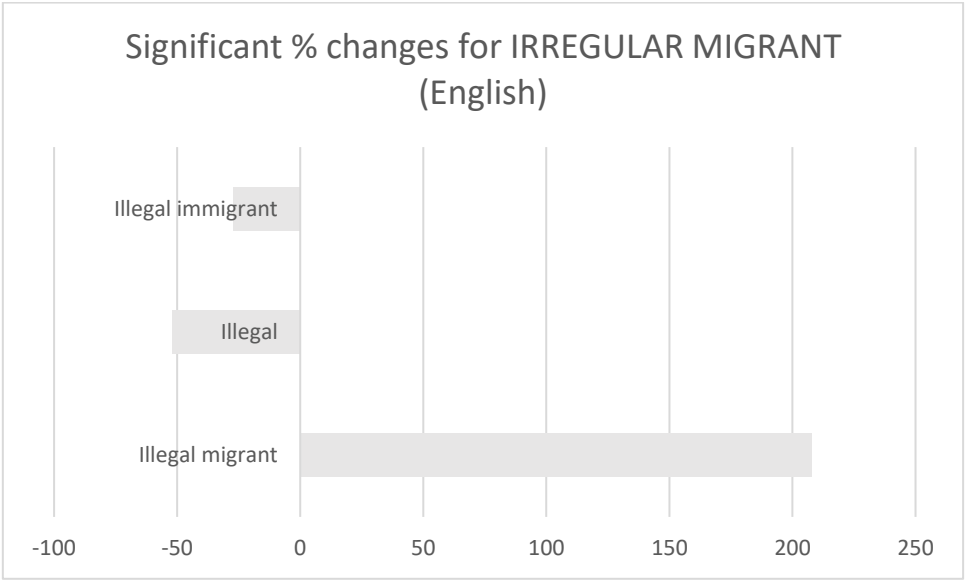


Figure 16: Percentage change in relative frequency of terms in conceptual category IRREGULAR MIGRANT for English news articles

In relation to the overall number of terms identified across all categories in each subcorpus, the conceptual category IRREGULAR MIGRANT saw a proportional decrease in the English part of the newspaper corpus, containing 16.04% of the total terms in the earlier subcorpus and only 10.67% of the total terms in the later subcorpus. As one of the central concepts in the field of migration, especially in the press, it is interesting to note this decrease in frequency of this concept in English newspapers over time despite the consistently prominent coverage of migration as a topic in the print media.

In the English conceptual category IRREGULAR MIGRANT, the term '*illegal entrant*', having appeared with a high relative frequency (354.22pmw) in the earlier subcorpus, is completely absent from the later subcorpus, constituting a very significant fall in frequency. This term was adopted in UK legislation (*Immigration Act 1971*) and is still occasionally used in legal and government documents (e.g. Home Office, 2022), but appears to have entirely fallen out of favour in the print media. Other terms in this category that recorded a significant fall in frequency were '*illegal immigrant*' (-27.28%) and '*illegal alien*', which is entirely absent from the later subcorpus. Meanwhile, the terms '*illegal migrant*', '*undocumented migrant*' and '*irregular immigrant*' all saw a statistically significant increase in frequency. Indeed, the latter two do not appear at all in the earlier subcorpus. Some of these terminological trends, namely the falls in frequency of numerous terms using the adjective '*illegal*' and the increase in the use of terms using alternative adjectives such as '*undocumented*' and '*irregular*', suggest an attempt from some newspapers to eradicate the criminalising nature of the terminology used in articles on the topic of migration. Indeed, the results for the English terminology for this category shows

certain terms carrying negative connotations falling out of favour while other more neutral or nuanced terms see a subsequent rise in frequency. In the earlier time frame, the only adjective used in this category is '*illegal*'. However, this changes in the later time period with the introduction of other adjectives, including '*irregular*', '*undocumented*' and '*unauthorised*'. This reflects a response to the many calls for changes to this terminology, as mentioned in section 4.3.3, arguing that only an action can be labelled illegal, not a person.

These results indicate the possible start of a trend towards using alternative adjectives, but in the large majority of cases, the adjective '*illegal*' persists. Indeed, although the relative frequency of the term '*illegal immigrant*' has fallen, its proportional frequency within the category IRREGULAR MIGRANT has actually increased, from 61% to 73%, due to the fall in the overall number of terms in this category. In the same vein as the proportional increase of '*illegal immigrant*', the term '*illegal*' also continues to be used as a nominalised adjective in the later subcorpus, as per the following example:

> *Caroline Nokes announced the fresh crackdown on* ***illegals***
>
> (Daily Mail, 2018).

Indeed, despite a modest decrease, this particularly pejorative usage persists in the later subcorpus.

In this case, therefore, there is a very mixed picture, with no uniform terminological trend emerging. The number of alternatives to the term '*illegal*' has clearly risen, but the use of '*illegal*' has also remained strong in the later subcorpus. Indeed, a range of complex and contradictory factors give rise to ambiguity with regard to English terminological trends in newspapers in the conceptual category IRREGULAR

MIGRANT. On one hand, there appears to have been efforts in some quarters to move away from terminology with negative connotations such as '*illegal entrant*' and '*illegal alien*' in favour of more neutral and less criminalising and dehumanising terminology, such as '*undocumented migrant*' and '*irregular migrant*'. Most research on UK media discourse in the field of migration points to an overwhelmingly negative portrayal of migrants, characterised by high levels of hostility (e.g. Berry, Blanco-Garcia and Moore, 2015), but these results do show some signs of an increased empathy towards the plight of migrants in the language used. On the other hand, other trends identified here demonstrate continued use in newspapers of terminology often deemed problematic, with notable increases in absolute frequency for terms such as '*illegal migrant*' and proportional increase for the term '*illegal immigrant*', which grows increasingly predominant within the category. These contradictory trends could be the result of the broad range of newspapers with divergent political stances contained in the *MigrationInTheNews Corpus*, as discussed in 3.3.1, the effect of which is examined in more detail in the discussion in section 5.5 below.

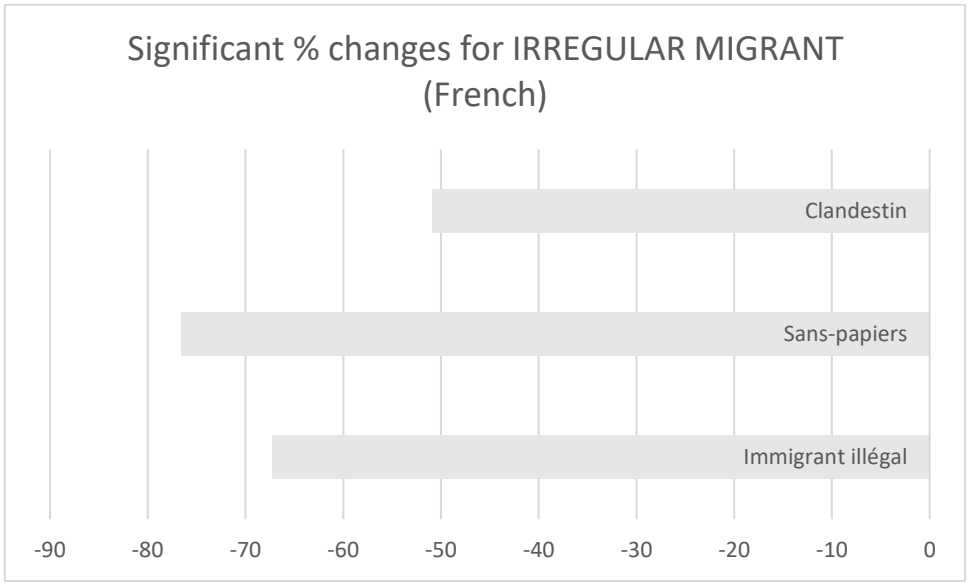


Figure 17: Percentage change in relative frequency of terms in conceptual category IRREGULAR MIGRANT for French news articles

In common with the observation of the English part of the corpus, the proportion of terms in the conceptual category IRREGULAR MIGRANT compared to the overall number of terms also falls significantly in French. The category contains 23.27% of all terms identified in the earlier subcorpus, compared to just 10.89% in the later one. This is an even more significant proportional decrease than in the English part, indicating a reduced focus on the concept of IRREGULAR MIGRANT in the French press over time despite the continued prominent coverage of migration as an issue.

As with the English part, there have also been significant changes in frequency for terminology used for the concept IRREGULAR MIGRANT in the French part of the corpus. The very high number of terms in this conceptual category (22) underlines the level of variation and instability around this concept. Indeed, 17 out of the 22 terms saw a statistically significant change in frequency over the time period. Eight new terms appear in the later subcorpus having been completely absent in the earlier subcorpus, the majority of which carry negative connotations, namely '*migrant illégal*' [illegal migrant], '*migrant clandestin*' [clandestine migrant], '*migrant sans papiers*' [migrant without papers], '*étranger entré illégalement*' [illegally entered foreigner] and '*immigrant sans papiers'* [immigrant without papers]. Taken together, these terms have increased from 0% of the conceptual category in the earlier subcorpus to 12% in the later one. Such terms are in direct opposition to the drive from some quarters, as mentioned in chapters 2 and 4, to improve terminology for this concept by moving away from directly labelling migrants as '*illegal*' or '*clandestine*' or defining their entire existence by way of their relationship to the country's immigration legislation. In total, the proportion of terms which use the adjectives '*illégal*' or '*clandestin*' increases from 55% in the earlier subcorpus to 59%

in the later one. Although it is true that a number of other terms with negative connotations in this conceptual category have either fallen out of usage completely, such as '*immigré clandestin*' [clandestine immigrant] and '*immigrant clandestin*' [clandestine immigrant], or, as shown in figure 17, fallen very significantly in frequency - '*clandestin*' [the clandestine] (-50.95%) and '*sans papiers*' [(person) without papers] (-76.59%) - the latter two terms still predominate in this category, representing 48% of the total terms. The below examples taken from the corpus illustrate how these terms are used:

> *Les experts soulignent cependant qu'ils ne prennent pas en compte les* ***clandestins*** *et que cette liste recouvre des réalités très diverses*.
>
> [*Experts stress, however, that they do not take* ***the clandestine*** *into account and that this list covers extremely varied realities*]
>
> (Le Figaro, 2015)

> *La police pourra se rendre sur le lieu de résidence des* ***sans papiers*** *pour exécuter une décision d'éloignement*.
>
> [*The police may go to the residence of the* ***[people] without papers*** *to enforce a removal decision*]
>
> (La Croix, 2015)

In contrast, as stated above, the alternative, less pejorative term '*migrant en situation irrégulière*' experienced a significant rise in relative frequency in this category, albeit from a base of 0 in the earlier subcorpus (from 0pmw to 17.36pmw). This is a term advocated by the International Organization for Migration for this concept, indicating

an effort from some newspapers to use less demonising terminology when discussing this topic. Moreover, the second most frequent term in the later subcorpus also uses the same construction, '*étranger en situation irrégulier*' (16% of the category), which also places the focus on the situation rather than the individual. Nonetheless, these terms taken together only equate to 18% of the total terms used to express IRREGULAR MIGRANT in the later subcorpus. Therefore, despite some minor signs of more empathetic language being used in this category, the general trend for French newspapers for the concept IRREGULAR MIGRANT therefore appears to be a persistence of terms which tend to vilify, denigrate and criminalise the migrants themselves.

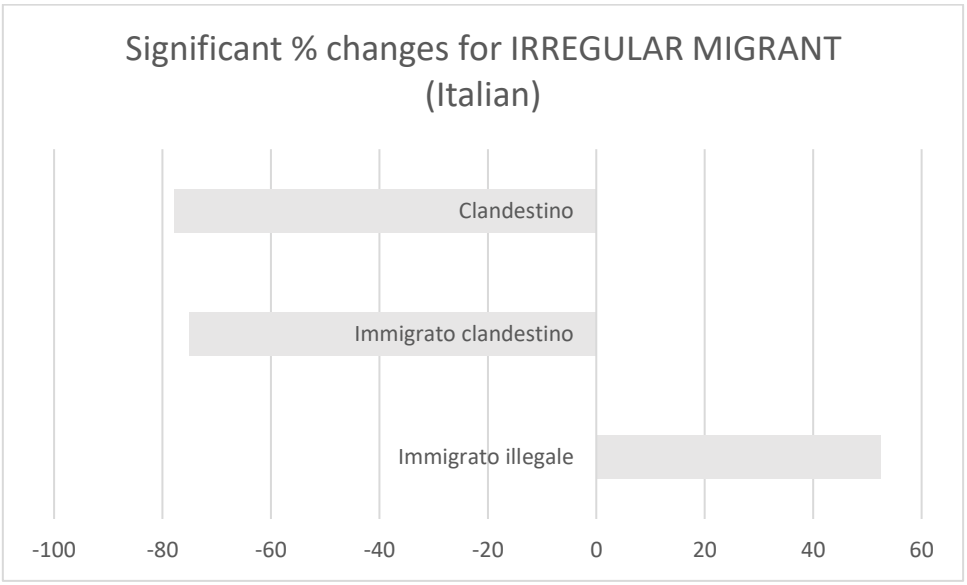


Figure 18: Percentage change in relative frequency of terms in conceptual category IRREGULAR MIGRANT for Italian news articles

As observed with both the English and French parts of the corpus, the proportion of the conceptual category IRREGULAR MIGRANT out of the overall number of migration terms identified has fallen significantly. For the Italian part of the earlier subcorpus, the category accounted for 20.68% of the overall terms, compared to just

7.49% in the later subcorpus. Consequently, it would appear that there has been a uniform drop off for this category across all three languages in terms of the focus accorded to it by the press in spite of the continued prominence of the topic of migration in the press throughout both time periods.

Also in line with the English and French results, the findings for the Italian part of the newspaper corpus demonstrate significant changes in the conceptual category of IRREGULAR MIGRANT. All but one of the terms in this category recorded a statistically significant change in frequency over the time period. The word '*irregolare*' has seen significant increases as both an adjective in multi-word terms and as a stand-alone nominalised adjective. '*Irregolare*' appears in only 3% of terms in the category during the earlier time period, compared to 33% in the later period. Specifically, having not appeared at all in the earlier subcorpus, the terms '*migrante irregolare*' [irregular migrant], '*straniero irregolare*' [irregular foreigner] and the nominalised adjective '*irregolare*' [the irregular] all recorded statistically significant increases. Meanwhile, the more pejorative term '*clandestino*' [clandestine] saw a fall in frequency, both as an adjective in multi-word terms such as '*immigrato clandestino*' [clandestine immigrant] (-75.02%), and when used as a nominalised adjective: '*clandestino*' [the clandestine] (-77.78%), as shown in figure 18 above. In fact, these are among the terms which recorded the most significant frequency drops across all conceptual categories for Italian. Indeed, the adjective '*clandestino*' appears in 92% of the total terms in the conceptual category in the earlier subcorpus, which falls to 53% in the later subcorpus.

While the use of the term '*irregolare*' appears to mark a progressive shift to a less pejorative label, the fact that it is used most frequently as a nominalised adjective, '*irregolare*' [the irregular], contradicts this somewhat. Indeed, although the

nominalisation of adjectives in this way is a common linguistic device in Italian, the use of the term '*irregolare*' as a noun in this case lacks the human quality it has when used as an adjective in combination with '*migrante*' (cf. Robinson, 2018). This is yet more apparent with another nominalised adjective in this conceptual category, '*illegale*' [the illegal], which also recorded a significant increase in frequency over the time period, rising from 0pmw to 28.34pmw. Indeed, despite the decrease in frequency of a number of pejorative terms in the category IRREGULAR MIGRANT, there does not appear to be a consistent terminological trend towards more humanising, empathic language to express this concept in Italian newspapers given the corresponding rise in nominalised adjectives '*irregolare*' and '*illegale*'. Examples taken from the corpus shows how these terms are used:

> *Resta aperta però la questione degli **irregolari** già presenti in Italia*
>
> [*The issue of **irregulars** already in Italy remains open*]
>
> (La Stampa, 1998)

> *Non si può dire solo che bisogna rimpatriare tutti gli **illegali**, perché sappiamo bene che ciò non può avvenire*.
>
> [*You can't just say we need to repatriate all the **illegals**, because we are well aware that can't be done*]
>
> (Il Giorno, 2017)

In summary, the conceptual category IRREGULAR MIGRANT has undergone substantial change in all three languages. For English, French and Italian, the

number of terms in this category has reduced significantly as a proportion of the overall number of migration terms identified in the corpus, suggesting an orientation away from this concept by the press. For each language, this is the category which contains the highest number of term variants, indicating the high level of variation and instability present in this category across the time period. Indeed, this is further emphasised by the fact that in each language several new terms appear in the later subcorpus which were entirely absent in the earlier one and other terms fall out of use completely having appeared in the earlier subcorpus.

The corpus analysis shows that certain types of terms in the conceptual category IRREGULAR MIGRANT have fallen significantly in usage across all three languages. Several terms in each language which could be deemed problematic or deprecated as they carry significant negative connotations have decreased in frequency over the time period. For English, these terms include '*illegal entrant*', '*illegal immigrant*' and '*illegal alien*', for French, '*immigrant clandestin*' [clandestine immigrant], '*immigré clandestin'* [clandestine immigrant]' and '*clandestin*' [the clandestine], and for Italian '*clandestino*' [the clandestine], '*immigrato clandestino'* [clandestine immigrant] and '*clandestino extracomunitario'* [the non-EU clandestine]. This points to an acknowledgement in all three languages of the unacceptability of certain terms and a shift away from the adjectives '*illegal*' and '*clandestine*'. However, this trend is only partial, as there are also examples of terms containing these adjectives which have actually increased over the time period, such as *'illegal migrant'*, '*migrant clandestin'* [clandestine migrant] and '*illegale'* [the illegal].

5.3.4 MIGRANT WORKER

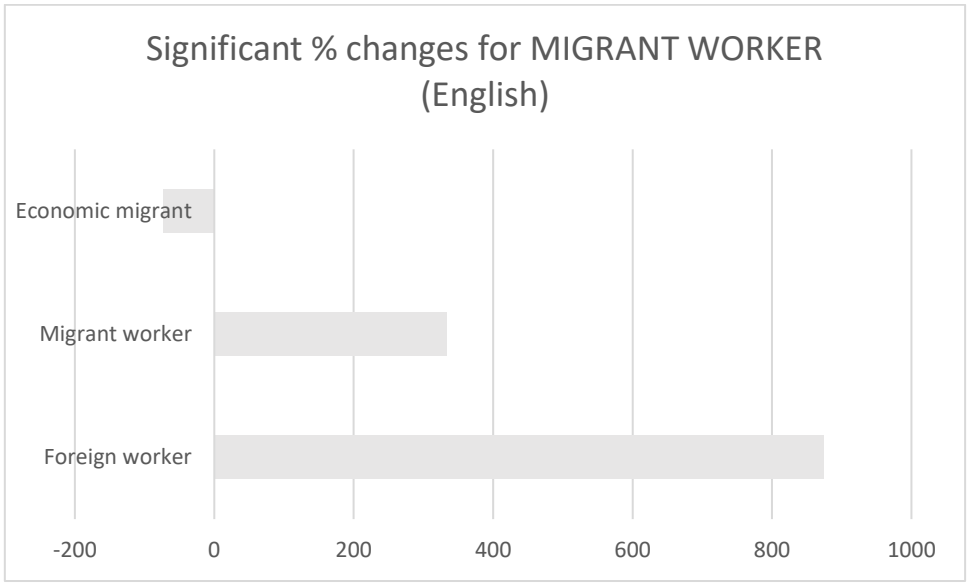


Figure 19: Percentage change in relative frequency of terms in conceptual category MIGRANT WORKER for English news articles

The results for the conceptual category MIGRANT WORKER in the English part of the corpus show very significant changes within this category over time, with statistically significant frequency changes for seven of the eight terms in the category. There are four new terms in the later corpus, which did not appear in the earlier one, along with two terms, '*migrant worker*' and '*foreign worker*', which experienced very substantial increases in frequency. With the exception of '*economic migrant*', all the significant changes are frequency increases, the result of which is that this category as a whole has significantly increased as a proportion of the overall migration-related terms identified in the corpus across the time period. It accounted for 1.48% of the total terms in the earlier subcorpus, which increases to 4.20% in the later one. The most frequently occurring term in the category in the earlier subcorpus was '*economic migrant*' before falling sharply as a proportion of the category by the later time period at the expense of '*foreign worker*', which is the

predominant term in the later subcorpus, representing 35% of the total terms. The proportion of the term '*migrant worker*' also rises (from 14% to 23%) across the time period.

The significant changes recorded in the conceptual category MIGRANT WORKER are likely to be closely related to the coverage of the 2016 Brexit referendum and its aftermath, which fall within the time period covered by the later subcorpus. Indeed, the increasing number of workers arriving from overseas, especially from Central and Eastern Europe, led to growing concern among the British public, with fears for the effects on the employment prospects for UK nationals and the potential for such immigration to drive down wages (Clarke, Goodwin & Whiteley, 2017, pp. 11-12). The topic of migrant workers was therefore a key theme during this time, as it was among the most salient issues in the debate leading up to the Brexit vote. This high level of coverage for this topic was true of all newspaper types (tabloid, mid-market and broadsheet) as per the below examples from the Daily Mail and the Financial Times:

> *British Firms where **Foreign Workers** are the Norm*
>
> (Daily Mail, 2016)

> *Immigration was already the hottest topic in Britain's EU debate even before the number of **migrant workers** from the bloc hit a new record.*
>
> (Financial Times, 2016)

It is also interesting to note that although the neutral term '*migrant worker*' saw a large increase in frequency, the term '*foreign worker*', with more negative

connotations, underwent an even more dramatic increase, becoming the predominant term in the category, surpassing '*migrant worker*' in the process. This terminological trend may also point to the growing demonisation in the UK press of migrant workers in the build-up to the referendum.

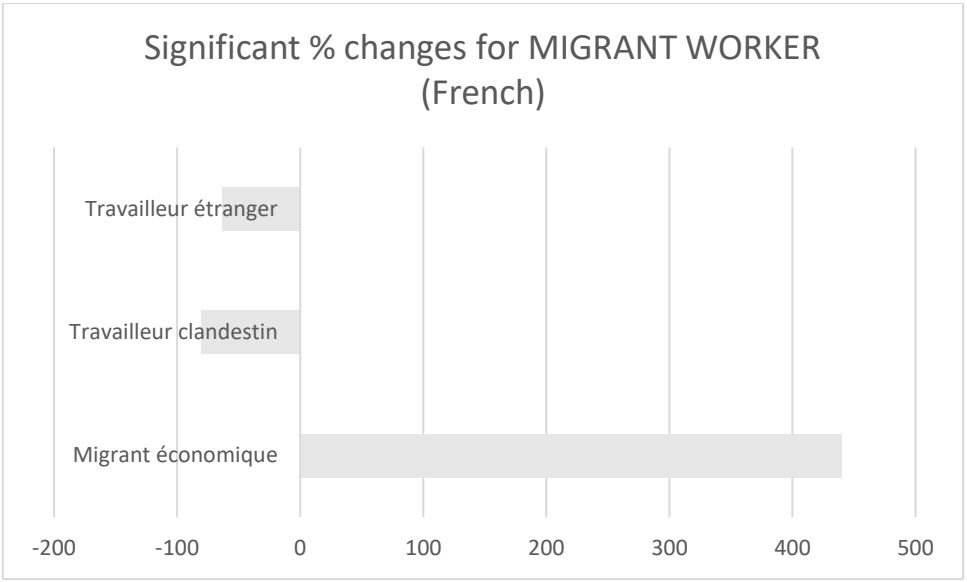


Figure 20: Percentage change in relative frequency of terms in conceptual category MIGRANT WORKER for French news articles

It is interesting to note that the category MIGRANT WORKER in the French part of the corpus fell in proportion to the overall number of terms over the time period. Having accounted for 2.18% of the overall terms in the earlier subcorpus, this fell to just 1.60% by the later one. This is in stark contrast to the substantial rise in prevalence of this concept recorded in the English part, suggesting a shift in focus away from this as a topic of discussion in the French press. The Brexit referendum was covered by French newspapers, as per the below example:

> *Le nombre total d'immigrants au Royaume-Uni est de 8.1 millions et la part des* ***travailleurs immigrés*** *sur le marché du travail a augmenté de 10% à 16% entre 2004 et 2014*
>
> [*The total number of immigrants in the UK is 8.1m and the proportion of* ***immigrant workers*** *on the labour market increased from 10% to 16% between 2004 and 2014*]
>
> (Les Echos, 2016)

However, the details of the debate understandably did not receive the same level of coverage as in the UK, and therefore the concept MIGRANT WORKER in general received less attention in France.

Within the category, six of the eight French terms recorded a statistically significant frequency change. Having not appeared at all in the earlier subcorpus, the term '*salarié étranger*' [foreign employee] rises to become the third most prevalent term in the category. The term '*travailleur immigré*' [immigrant worker] remains the most common term in the category despite a fall in relative frequency down from 39% to 24%, and is joined by '*migrant économique'* [economic migrant], also on 24% in the later subcorpus. Moreover, taken together, the terms '*travailleur migrant*' [migrant worker] and '*travailleur étranger*' [foreign worker] account for only 17% of the total terms in the category by the later subcorpus, while their English equivalents represent 58% of the category for English, underlining the significant differences between the two languages for this concept.

Similar to the English part, there is a proportional increase in the Italian part for the category MIGRANT WORKER, which rises from 1.08% to 1.66% of the overall

number of terms identified[15]. There are five terms in the category in total, but only one term, '*lavoratore straniero*' [foreign worker], appears in both, as per the following example:

> *Il 10% della produzione economica nazionale è legato all'impiego di* ***lavoratori stranieri***
>
> [*10% of national economic production is connected to the employment of* ***foreign workers***]
>
> (La Nazione, 2018)

The other two terms in the category in the earlier subcorpus, '*lavoratore extracomunitario*' [non-EU worker] and '*lavoratore immigrato*' [immigrant worker] are entirely absent from the later subcorpus. Meanwhile, having been absent from the earlier subcorpus, the overwhelmingly most frequent term in this category in the later time period is '*migrante economico*' [economic migrant], which represents 66% of the total terms by that time, as per the example given below:

> *I vantaggi per i Paesi ospitanti sono facili da vedere: i* ***migranti economici*** *spesso lavorano di più per meno soldi rispetto alla gente del posto*
>
> [*The benefits for host countries are plain to see:* ***economic migrants*** *often work longer hours for less money compared to local people*]
>
> (La Stampa, 2015)

As this example shows, the term '*migrante economico*' is often used to frame migrants as outsiders willingly depriving local people of job opportunities. The

[15] No graph is included for MIGRANT WORKER for Italian since the only terms showing a statistically significant difference in frequency all have a value of 0 in either the earlier or later time period, thus percentage change values are unrepresentative in this case.

particularly high unemployment in Italy during this period, peaking at around 12% (World Bank, 2023), resulted in migrant workers often being unfairly blamed in the press and by politicians for exacerbating the job shortages in the country. Since the term '*migrante economico'* specifically depicts migrant workers from a bureaucratic, economic perspective, this may at least in part explain its dramatic rise during this time period in Italian print media, as it conveniently serves the purpose of those news outlets seeking to accentuate this particular aspect.

In sum, the category MIGRANT WORKER undergoes substantial terminological changes in all three languages over the time period, but these changes are manifested in very different ways across the languages. Indeed, this concept sees a very significant overall rise in the English part, combined with substantial changes in frequency for the terms within the category. This trend is likely to be largely attributable to the Brexit referendum, and the frequency of coverage given to the theme of migrant workers in the UK. The proportion of this category also rises in Italian, while its terminological profile changes radically, with '*migrante economico'* [economic migrant] becoming the most prevalent term by a substantial margin, overtaking '*lavoratore straniero'* [foreign worker] in the process. French, on the other hand, sees a decrease in the proportional representation of this concept overall, with the term '*travailleur étranger'* [foreign worker] becoming the most popular term in French news articles by the later time period.

5.3.5 THIRD-COUNTRY NATIONAL

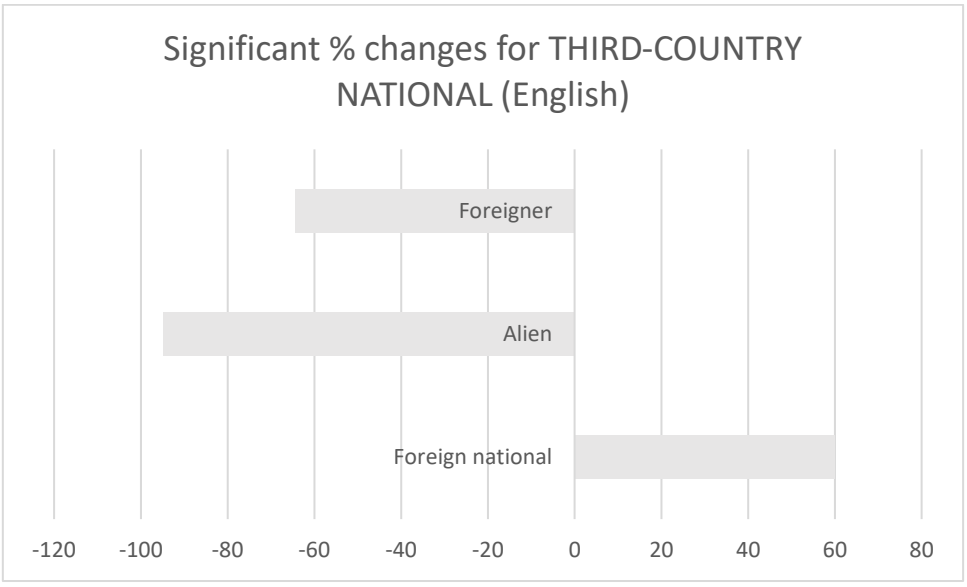


Figure 21: Percentage change in relative frequency of terms in conceptual category THIRD-COUNTRY NATIONAL for English news articles

The conceptual category THIRD-COUNTRY NATIONAL accounts for 8.12% of the overall number of terms in the English part of the earlier subcorpus. By the later subcorpus, this proportion drops to just 3.66%. This decrease indicates a reduced reference to this particular concept in UK newspapers over the time period being studied. The results for this category in English show very significant changes in the terms used to express the concept THIRD-COUNTRY NATIONAL. Indeed, all ten terms in this category show a significant difference in frequency over the time period.

It should be noted that, while the EU definition of the concept THIRD-COUNTRY NATIONAL specifically refers to a person from a country which is not an EU Member State, news articles in all the languages in question are often less rigorous in differentiating between people from within the EU and those from outside it. For this reason, it is not possible in many cases in the *MigrationInTheNews Corpus* to know

whether a term refers to EU nationals or non-EU nationals, or indeed, a combination of both, even taking the context of the article into account. This is illustrated by the below examples:

> *Those **foreigners** are vital to the workforce, particularly when so many employers report difficulty in finding the right people to fill a role*
>
> (The Evening Standard, 2015)

> *Syndicats et associations dénoncent une logique en totale contradiction avec l'accueil et l'intégration des **personnes étrangères***
>
> [*Unions and charities condemn an approach that is in complete contradiction with the welcoming and integration of **foreign persons***]
>
> (L'Humanité, 2016)

Whether the use of the terms in the above cases refers to non-EU citizens is unclear and the distinction is not made in the articles in question, which is a common trait of many articles in all three languages and in both time periods.

In any case, there are substantial decreases in English of the frequency of the terms '*foreigner*' and '*alien*' (-64.50% and -94.87%), as shown in figure 21 above, which both have clearly negative semantic prosody. Taken together, these two terms constituted 89% of the total terms used in the category in the earlier time period. By the later time period, the proportion of the term '*foreigner*' falls from 57% to 50%, while '*alien*' decreases even more dramatically, from 32% to just 4%. Conversely, the more neutral term '*foreign national*' saw a significant increase in relative frequency, with a percentage increase of 59.98%, amounting to a proportional

increase in the category from 4% to 16%, used in sentences such as the following, taken from the corpus:

> *…three-quarters of the growth in employment over the last year was accounted for by* ***foreign nationals***
>
> (The Guardian, 2015)

Likewise, the five new terms which only appear in the later subcorpus *(‘foreign-born person’, ‘person from overseas’, ‘foreign citizen’, ‘non-European*’ and ‘*non-EU citizen*’) all have neutral connotations and account for 29% of the total terms in the later subcorpus. Based on these results, there appears to be a clear shift in the UK press towards more neutral terms to express those from countries outside the UK or EU. Often criticised for the use of dehumanising labels, as highlighted in chapter 2 (section 2.3), it is interesting to note the newspapers apparently shifting away from those terms which have the effect of othering those to which they are applied, casting them as outsiders. Indeed, the term ‘*outsider*’ itself also significantly falls in frequency in the later subcorpus. However, it is important to note that the term ‘*foreigner*’ still accounts for 50% of the total terms in this category, and that this category only represents 3.66% of the total number of migration-related terms in the later subcorpus. With this in mind, this shift in the UK press from negative to more neutral terms to express the concept of THIRD-COUNTRY NATIONAL should not be overstated.

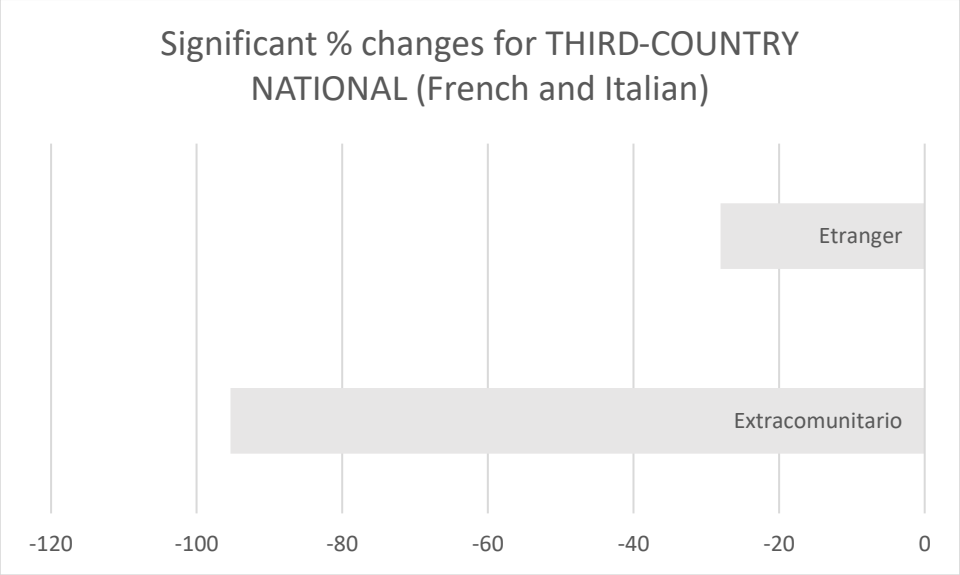


Figure 22: Percentage change in relative frequency of terms in conceptual category THIRD-COUNTRY NATIONAL for French and Italian news articles

In the French part of the corpus, the conceptual category THIRD-COUNTRY NATIONAL occupies a much larger proportion of the total number of terms than in English. Indeed, in the earlier subcorpus, this category accounts for 30.00% of the overall number of terms. Like in English, this proportion decreases by the later subcorpus, falling to 19.16%, but nonetheless remains much higher. This suggests a much greater focus on this particular concept in French print media than in English.

Although six of the eight terms in this category undergo a significant change in frequency between the time periods, the overwhelmingly predominant term in both time periods is the general and somewhat negative term '*étranger*' [foreigner]. Its fall in relative frequency, as shown by figure 22 above, is largely the result of the overall decrease in the prevalence of this concept in the later subcorpus. Indeed, despite the fall in relative frequency, it still represents 94% of the total terms in the category in the later time period. Four new terms appear in the later subcorpus, having been

absent in the earlier one, but taken together only constitute 4% of the overall terms in this category in the later subcorpus. Consequently, there appears to be high general stability in this category for French with no major shifts in terminological direction between the time periods.

With regard to Italian, the conceptual category THIRD-COUNTRY NATIONAL mirrors the English and French parts insofar as the category accounts for a smaller proportion of the overall terms in the later subcorpus compared to the earlier one. Indeed, in the Italian part of the earlier subcorpus, the category accounts for 26.22% of the overall terms identified, and only 12.64% in the later subcorpus. In Italian, there is little significant change in this category over time with regard to the frequency of the terms within it. Indeed, only two terms, the now pejorative nominalised adjective '*extracomunitario*' [the non-EU] and '*non europeo'* [non-European], experience a significant frequency change, both decreases. '*Non europeo'* does not appear at all in the later time period, while '*extracomunitario*' records a very dramatic fall in frequency with a percentage change value of -95.34%, as shown in figure 22. An example of the use of this term in the earlier subcorpus is as follows:

> *È evidente che i nostri servizi segreti sono preoccupatissimi per i flussi di* ***extracomunitari***
>
> [*It is clear that our secret services are extremely worried about the flows of* ***non-EU migrants***]
>
> (La Stampa, 1998)

The change in the connotations attached to the term largely explain this substantial decrease in frequency. Initially, the term '*extracomunitario*' was used in neutral

Italian legal texts to describe any person who is not an EU citizen (ISMU, 2006). However, over time, the term has acquired very strong negative overtones of 'otherness' in Italian society (Russo Bullaro, 2010). Indeed, '*extracomunitario*' is now usually used to express hostility towards migrants and tends to be applied selectively to refer solely to those coming from poorer countries (Treccani, 2022). The fall in the newspapers' usage of this term therefore indicates that it has drastically fallen out of favour among almost all parts of the Italian press regardless of their political stance as a way of expressing the concept of THIRD-COUNTRY NATIONAL. As a result of this change, although the term '*straniero*' [foreigner] undergoes a slight decrease in relative frequency across the time period, it constitutes a much larger proportion within the category, rising from 50% in the earlier subcorpus to 88% in the later one. Therefore, by the later time period, '*straniero*' is clearly the predominant term within this category.

In sum, for the conceptual category THIRD-COUNTRY NATIONAL, these results show that there is significantly more variation and instability in English than in French and Italian. In Italian, only two of five terms see a significant change in frequency, while in French, several new terms appear in the later subcorpus, but changes in both relative frequency and proportion of terms within the category remain relatively minor. In English, however, all terms in the category undergo statistically significant increases or decreases in frequency. As a result, the predominant English term in the category, '*foreigner*', accounts for only 50% of the overall terms in the category, while the equivalents in French and Italian, '*étranger*' and '*straniero*', account for 94% and 88% of the total category in the later time period. Moreover, English is also the language in which this category occupies the lowest proportion of overall migration terms identified in both the earlier and later subcorpora (8.12% and

3.66%). This is in stark contrast to the French and Italian parts of the corpus, where the category THIRD-COUNTRY NATIONAL represents a very substantial 30.00% and 26.22% of the overall terms in the earlier period respectively. While these figures drop to 19.16% and 12.64% in the later subcorpus, they are still much higher than for English.

5.3.6 UNION CITIZEN

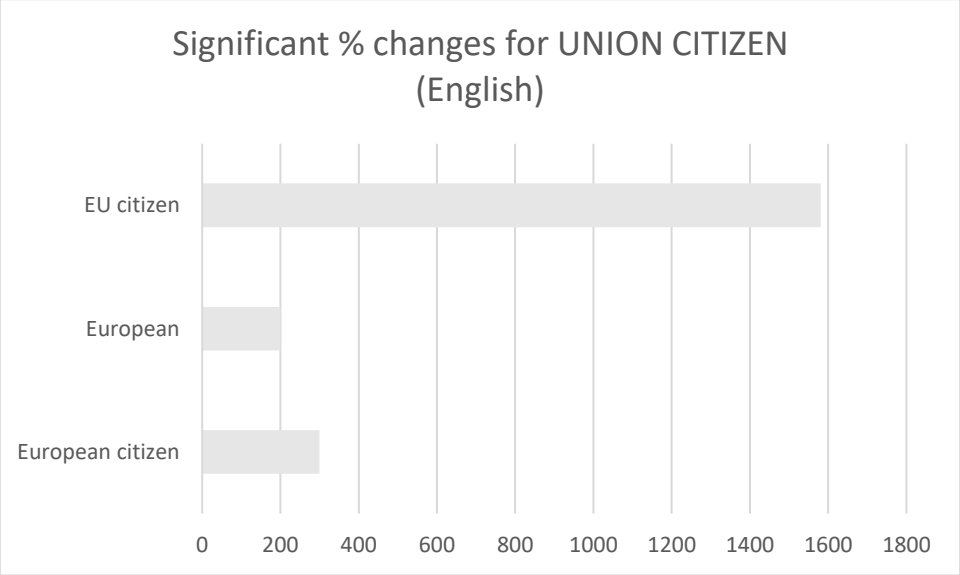


Figure 23: Percentage change in relative frequency of terms in conceptual category UNION CITIZEN for English news articles

In the English part, the concept UNION CITIZEN is another example of a category which saw a very significant increase as a proportion of the overall number of terms. Having made up only 1.18% of the total terms in the earlier subcorpus, this concept increases to 7.68% in the later time period. All four terms in the English category saw a statistically significant increase in frequency. The terms '*EU citizen*' (1,581.15%),

‘*European*’ (199.89%) and ‘*European citizen*’ (299.81%) experience huge increases in frequency. The term ‘*EU citizen*’ appears to become established as the most common term for this concept by the later subcorpus, accounting for 68% of the overall terms in the category, having surpassed the broader term ‘*European*’ over the time period. In common with the category MIGRANT WORKER, in section 5.3.4, it appears that the general topic of EU citizenship is discussed in UK newspapers much more frequently over time. It is interesting to note in this regard that the notion of EU citizenship was enshrined in the Maastricht Treaty, signed in 1993 (Maas, 2007, pp. 1), and yet terms to express the concept of UNION CITIZEN appear relatively rarely in newspapers around this period. This suggests therefore that it took some time for this concept to become embedded in press discourse, but by the time of the later subcorpus, it is referenced relatively frequently. In addition, it is also likely that discussions around the Brexit referendum also contributed to this rise in prevalence of this concept, as per the below example:

> *She [The Home Secretary] said that while it was “unthinkable” that* ***EU citizens*** *would be asked to leave the UK after Brexit, she could not offer them any guarantees about their immigration status while negotiations continue.*
>
> (Financial Times, 2017)

Indeed, during this period, the status of union citizens living and working in the UK was hotly debated both during and after the vote, and became a central issue of the post-referendum negotiations.

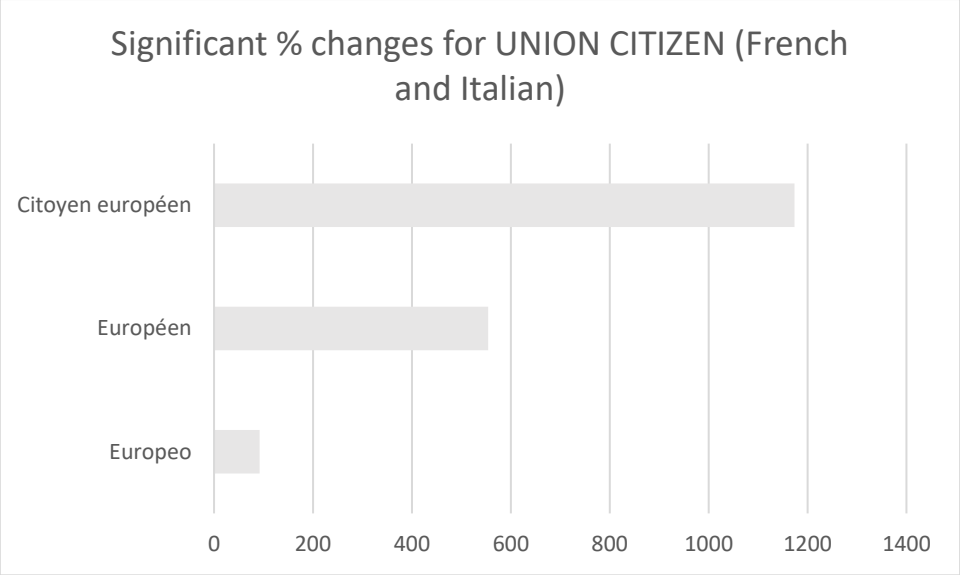


Figure 24: Percentage change in relative frequency of terms in conceptual category UNION CITIZEN for French and Italian news articles

There is also a proportional increase in the prevalence of this concept in the French part of the corpus. It accounts for just 0.89% of the total terms in the earlier subcorpus, increasing to 4.51% by the later time period. This indicates that the concept UNION CITIZEN has also increased in coverage to a similar extent in the French press. Within the category, four of the six terms undergo a statistically significant change in frequency, most markedly '*citoyen européen'* [European citizen] and '*Européen'* [European], which both record very substantial percentage changes (1,173.91% and 554.05%). In contrast to English, it is the term '*Européen*' that has become predominant in this category for French, accounting for 83% of all terms. This overall increase in terms can also be attributed in part to discussions around the fate of union citizens living in the UK, as per the below example:

> *Le Royaume-Uni n'obtiendra pas l'accès au marché intérieur s'il n'accepte pas les quatre libertés de circulation, y compris pour les* ***Européens*** *qui comptent s'installer et travailler outre-Manche*
>
> [*The UK will not get access to the internal market if it does not accept the four freedoms of movement, including for* ***Europeans*** *looking to settle and work there*]
>
> (Le Figaro, 2016)

In addition, two terms appear in the later subcorpus having been absent in the earlier one ('*ressortissant européen*' [European national] and '*citoyen de l'UE*' [EU citizen]) while two other terms fall out of usage in the later subcorpus having appeared in the earlier one ('*ressortissant de l'union européen*' [national of the European Union] and '*ressortissant communitaire*' [community national]), meaning that the overall number of terms in the category remains at four.

Reflecting the results obtained from both the English and French parts of the corpus, the proportion of the overall number of terms constituted by the concept UNION CITIZEN also increases for the Italian part over time. This category accounts for 1.25% in the earlier subcorpus and rises to 2.66% in the later one. Within the category, there is a very substantial degree of variation, with only one term ('*Europeo*') out of the total seven appearing in both the earlier and later subcorpora. Although the term '*Europeo*' increases in relative frequency, it decreases as a proportion of the category as a whole with the introduction of several other terms to the category. The most frequently occurring of the new terms appearing only in the later subcorpus is '*cittadino europeo*' [European citizen], which accounts for 15% of the total terms in this category. The other new terms are '*cittadino UE*' [EU citizen],

‘*cittadino comunitario*’ [community citizen], ‘*cittadino dell’UE*’ [citizen of the EU] and ‘*cittadino dell’unione Europea*’ [citizen of the European Union]. The proliferation in terms over the time period in Italian indicates that rather than a gradual standardisation of terminology for this concept, there is in fact greater instability and variation, with six term variants appearing in the later subcorpus, compared to just two in the earlier one.

Consequently, the conceptual category UNION CITIZEN undergoes consistent increases across the three languages as a proportion of the overall terms. In English, there are significant increases for the terms ‘*EU citizen*’, ‘*European*’ and ‘*European citizen*’, in French for the terms ‘*Européen*’ [European], ‘*citoyen européen*’ [European citizen] and ‘*ressortissant européen*’ [European national], and in Italian for the terms ‘*cittadino europeo*’ [European citizen], ‘*cittadino comunitario*’ [community citizen] and ‘*cittadino UE*’ [EU citizen]. All three languages therefore show a clear rise in the prevalence of the concept of EU citizenship in press articles.

The overall number of occurrences of terms expressing the concept of UNION CITIZEN across the three languages combined in the earlier corpus is 83 (39 for English, 20 for French and 24 for Italian). This number increases dramatically by the later subcorpus to 286 (206 for English, 119 for French and 61 for Italian). This indicates that the notion of EU citizenship has become a familiar concept in all three languages, frequently used in the non-specialist press, especially in the context of the UK Brexit referendum. However, it is interesting to note the continued instability around the terminology for this concept in each language. Indeed, the number of term variants in the conceptual category UNION CITIZEN for each language does not show any indication of decreasing over time. For English, the category contains three term variants in both the earlier and later subcorpora, while for French, there

are four term variants in both time periods. Even more starkly, the number of Italian term variants in the category actually increases from two to six. Taken together, the results for the conceptual category UNION CITIZEN show that the concept has increased significantly in prevalence in each language, but that there is substantial variation within each language regarding the terminology used by the press to express this concept.

5.4 Discussion

The results and analysis set out in this chapter bring to light a number of intralingual and interlingual trends regarding migration terminology used in the press in English, French and Italian. Firstly, the overall results indicate large-scale change across the time period within each language regarding the frequency of the terms identified, with English showing the most substantial change (88% of terms displaying a statistically significant change in relative frequency), followed by Italian (78%) and French (72%). There were more statistically significant increases than decreases across all three languages. There is, however, a marked difference in the total numbers of terms identified in each language across the time period. While the number of terms identified for English fell significantly, the number of French terms rose significantly and the number of Italian terms remained stable, indicating a substantial difference in terminological dynamics among the languages.

With regard to the individual conceptual categories as a proportion of the overall terms, a number of different trends emerge from this analysis. Certain conceptual categories display uniform behaviour across the three languages. Indeed, the categories MIGRANT and UNION CITIZEN record a significant increase as a

proportion of the total terms in English, French and Italian. Likewise, there is a uniform decrease in the prevalence of the categories IRREGULAR MIGRANT and THIRD-COUNTRY NATIONAL across the three languages. Meanwhile, there are other conceptual categories which display starkly different trends across the three languages, including ASYLUM SEEKER, REFUGEE and MIGRANT WORKER. There are also a number of conceptual categories which only represent over 1% of the overall number of terms in one language, as shown in table 18, namely DISPLACED PERSON in Italian, REJECTED APPLICANT FOR INTERNATIONAL PROTECTION in French, and INTERNATIONAL STUDENT in English. The other categories, representing under 1% of the overall terms, do not appear with sufficient frequency to enable clear trends to be identified.

Interesting trends are also evident within the individual conceptual categories. There is some indication of an acknowledgement in some areas of the print media of the importance of accurate terminology and avoiding terms with clearly negative connotations, hence the significant increase in terms such as '*foreign national*' and '*undocumented migrant*' in English, '*personne née à l'étranger*' [person born abroad] and '*migrant en situation irrégulière*' [migrant in an irregular situation] in French and '*migrante irregolare*' [irregular migrant] in Italian, combined with concurrent decreases in demonising terms, including '*illegal immigrant'* and '*alien*' in English, '*étranger*' [foreigner] and '*clandestin*' [the clandestine] in French, and '*extracomunitario*' [the non-EU] and '*immigrato clandestino*' [clandestine immigrant] in Italian. However, this trend is somewhat tempered by the fact that other terms with negative connotations have persisted and, and in some cases, even increased in frequency in all three languages, including '*immigrato illegale*' [illegal immigrant], '*migrant clandestin*' [clandestine migrant] and '*illegal migrant*'.

The trends observed here are often attributable, at least in part, to particularly consequential and protracted news stories. These include, among others, the Brexit referendum, the situation of Syrian refugees or ongoing EU integration, leading, for instance, to especially significant increases in the prevalence of the category MIGRANT WORKER in English, ASYLUM SEEKER in French and Italian, and UNION CITIZEN in all three languages. However, the differences in the findings across the languages show that there is not always a direct correlation between the events occurring in a country or region and the terminology used in news articles. Indeed, the concepts addressed and the terms used for these concepts are also influenced by editorial choices regarding the focus and framing of news stories on the topic of migration. Framing refers to the way that a communication source presents and defines an issue with an aim to shape individual and societal attitudes towards that issue (de Vreese, 2005). This is apparent, for example, in the marked decrease in the frequency of the concepts REFUGEE and ASYLUM SEEKER in English, coupled with the concurrent increase in the frequency of the category MIGRANT, indicating an effort in the UK print media to shift the frame away from the plight of people fleeing persecution towards purportedly ‘less deserving’ migrants. After all, it is the case that journalists do not simply report facts, but instead write stories that have a structure and direction as well as, most critically, a viewpoint (Bell, 1991, pp. 147). Consequently, the terminology in the *MigrationInTheNews Corpus* is arrived at both due to the current events occurring the world but also as a result of the lens through which these events are seen and subsequently projected by the range of different news outlets featured in the corpus.

Connected to this, another key factor contributing to the differences in the findings between the languages in the analysis of the *MigrationInTheNews Corpus* is the

significantly different written press landscapes in the UK, France and Italy. Indeed, as examined in chapter 3, the print media in the UK has a much higher proportion of tabloid newspapers with sensationalist and populist perspectives, in comparison with France and Italy. Previous research has highlighted the proliferation of overtly negative terms in tabloid coverage compared to news articles in broadsheets (e.g. Taylor, 2014). This particularly partisan stance taken by the newspapers with the highest circulation in the UK (e.g. *The Sun, The Daily Mail*) is in stark contrast to the more neutral accounts in the broadsheet newspapers (e.g *Financial Times, The Independent*), which partially explains the significant level of variation detected in migration terminology across the English news articles. This tendency to interpret news in the UK print media as opinions is also reflected by the fact that only 15% of UK citizens tend to trust the written press, compared to 48% in France and 51% in Italy (European Union, 2019). Nevertheless, the above corpus analysis also shows contradictory trends in French and Italian terminology used in newspaper texts. Although the media landscape of these nations is not tabloid-dominated in the same way as the UK, there is nonetheless still a broad spectrum of political stances represented by the French and Italian newspapers included in the corpus (as set out in chapter 3), which explains the diversity of terminological trends identified for the range of concepts in migration explored above. As stated above, a detailed investigation into the specific terminological differences of these various newspaper types constitutes an interesting avenue for further study but falls outside the scope of this particular research project.

Chapters 4 and 5 have set out a comprehensive analysis of changes in migration terminology in EU legal texts and press articles in the three languages being investigated in this thesis. The following chapter constitutes the final stage of the

analysis process, combining data obtained from both corpora in order to draw conclusions regarding extent of influence of translation-induced language change on migration terminology.

6. Stage 3: Terminology comparison between the *EUMigrLaw Corpus* and the *MigrationInTheNews Corpus*

The objective of this chapter is to compare both synchronically and diachronically the migration terminology used in EU texts and print media in English, French and Italian with a view to identifying similarities and differences across the two different text types. The results of this analysis can therefore provide an insight on the extent of influence migration terminology in EU law may have on terms in the written press. This chapter examines the general terminological trends across the three languages in both corpora before looking in more detail at specific conceptual categories and the terms contained within them.

6.1 Methodology

Firstly, the terms in each conceptual category from the *EUMigrLaw Corpus*, sorted as per the methodological approach set out in chapter 3, are placed into a results table alongside the terms for each category from the *MigrationInTheNews Corpus*. These results are separated by language and time period, providing six results tables in total (two for each language to cover both time periods). These results tables also record the relative and absolute frequency of each term in both corpora. The absolute frequencies of all terms in a given conceptual category can be combined in order to calculate the prevalence of that category as a proportion of all categories in that time period. The proportion that each category represents in each subcorpus offers an insight into which topics were most prevalent in each text type,

language and time period, and the results of these calculations are set out in section 6.2.

Subsequently, the relative frequency values of each term can be used to calculate the percentage difference in frequency of each term between the two text types. These percentage differences can be calculated as follows:

First value – Second value/(First value + Second value)/2 x 100 = Percentage difference

It should be noted that this differs from the methodology used in the previous chapter, which calculated percentage *change* in order to determine diachronic terminological shifts. This aspect of the methodological approach in this chapter focuses on the extent of synchronic terminological difference between the two text types in question, hence the use of percentage *difference* in this case. As in the previous chapter, the percentage difference values for each term can be subjected to statistical significance testing using the log likelihood value, calculated in exactly the same way as described in chapter 5 (section 5.1). The p-values derived from the log likelihood calculations can again indicate whether a frequency difference of a term in two different corpora can be deemed statistically significant.

In addition to the above data, it is also useful to determine the prevalence of each term as a proportion of the total number of terms within that conceptual category. As per the methodology set out in chapter 5 (section 5.1), this provides complementary information on the terminological dynamics within a conceptual category in both text types. Table 19 below shows an example of the results for the English conceptual category MIGRANT WORKER for the earlier time period containing all the data described above for each term in this category. The percentage difference column is

highlighted in bold where the difference can be deemed statistically significant according to the p-value.

Table 19: Comparative results table for conceptual category MIGRANT WORKER (Eng 1992-1998)

TERM	Frequency in *EUMigrLaw* (absolute/relative)	As a proportion of conceptual category MIGRANT WORKER in *EUMigrLaw*	Frequency in *MigrationInTheNews* (absolute/relative)	As a proportion of conceptual category MIGRANT WORKER in *MigrationInTheNews*	Percentage difference
Immigrant worker	4/11.22	11.76%	8/20.99	18.18%	60.66%
Migrant worker	22/61.71	64.71%	6/15.74	13.63%	**-118.71%**
Foreign worker	5/14.03	14.71%	4/10.5	9.09%	-28.78%
Economic migrant	3/8.42	8.82%	26/68.22	59.09%	**156.05%**

The information contained in the results tables shows the extent of terminological difference or similarity between the text types during a given time period in each language. Thereafter, the results from both time periods can be analysed together in order to determine the extent of terminological convergence or divergence that has occurred between the two text types in each language over the time period. Indeed, this subsequent stage involves a process to produce three complementary indicators of terminological convergence/divergence to show whether the terminology used in the two text types in each language has grown more similar or become more different over time, the results of which are set out in section 6.3. These three indicators, explained in more detail below, provide values regarding i) frequency difference, ii) shared predominant terms, and iii) terms appearing in only one of the corpora.

The first indicator is the overall percentage of terms in each time period which have a statistically significant frequency difference across the two text types. An increase in

this percentage over the time period would indicate divergence of terminology, while a decrease would indicate convergence. The statistical significance of frequency differences can be calculated as per the previous chapter, where log likelihood > 3.84 = $p<0.05$.

The second indicator of terminological divergence/convergence focuses on the predominant term in each conceptual category. This indicator shows the percentage of concepts for which both text types have the same predominant term. Table 20 below provides an example for the conceptual category MIGRANT taken from the French 2015-2018 subcorpus. In this case, both text types have the same predominant term (highlighted in bold). Rather than taking relative frequencies into account, this indicator solely examines those terms which have, to a greater or lesser extent, become the most prevalent within that particular category.

Table 20: Example of category with same predominant term

	Term	**Proportion of category in EU texts**	**Proportion of category in news articles**
MIGRANT	***Migrant***	**94.82%**	**53.18%**
	Personne migrante	1.04%	0
	Expatrié	0	0.27%
	Immigré	1.24%	34.45%
	Immigrant	1.66%	9.55%
	Nouvel arrivant	1.24%	2.55%

The third indicator examines the proportion of terms which only appear in one of the corpora and is entirely absent from the other. For example, the English term '*displaced person*' appears with a relative frequency of 20.67pmw in the later time period of the EU corpus but does not appear at all in the newspaper corpus in the

same time period. This indicator therefore provides a percentage of terms which are not shared by both corpora. An increase in this percentage over time would point to a terminological divergence since more terms would be unique to one or the other corpus, while a decrease would indicate convergence.

Analysing these three indicators in combination offers a rounded assessment of whether the terminology used in the two text types across the three languages is becoming more similar or growing more different, the results of which are set out in section 6.3. This quantitative stage is then followed by a qualitative analysis to gain a deeper insight into the terminological dynamics present across the text types over the time period.

6.2 Quantitative results on proportion of concepts per subcorpora

Figures 25, 26 and 27 below show the percentage represented by each concept as a proportion of the total number of terms in each subcorpora. These figures show those concepts which represent at least 1% or more of the overall proportion in the majority of subcorpora. For reasons of length, the full results table showing all categories and the overall proportion they represent is included in appendix 3.

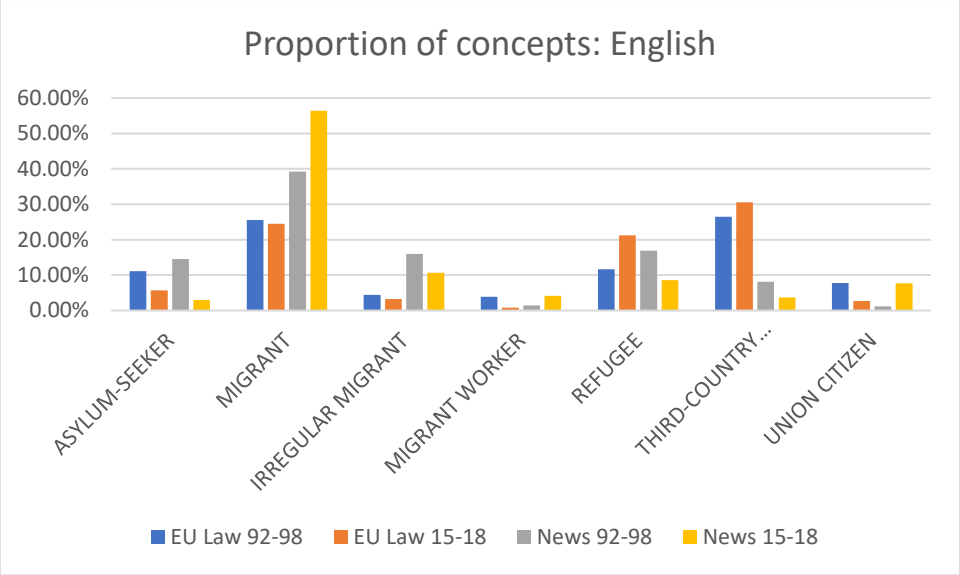


Figure 25: Each concept in each subcorpora as an overall proportion of English terms

The data for English, as illustrated in figure 25 above, shows two concepts which exhibit similar trends in both EU legal texts and news articles over the time period: ASYLUM-SEEKER and IRREGULAR MIGRANT. Both text types demonstrate a marked decrease in both concepts as a proportion of the overall number of terms. The five other concepts taken into account here all show significantly different trends in the two text types, with the concepts MIGRANT, MIGRANT WORKER and UNION CITIZEN falling in EU texts and increasing in newspapers articles. Conversely, the concepts REFUGEE and THIRD-COUNTRY NATIONAL increase in EU texts while decreasing in newspaper articles. Prior to examining the specific terms used in each conceptual category, these results suggest an overall divergence in the general concepts addressed by these two different English text types over the time period.

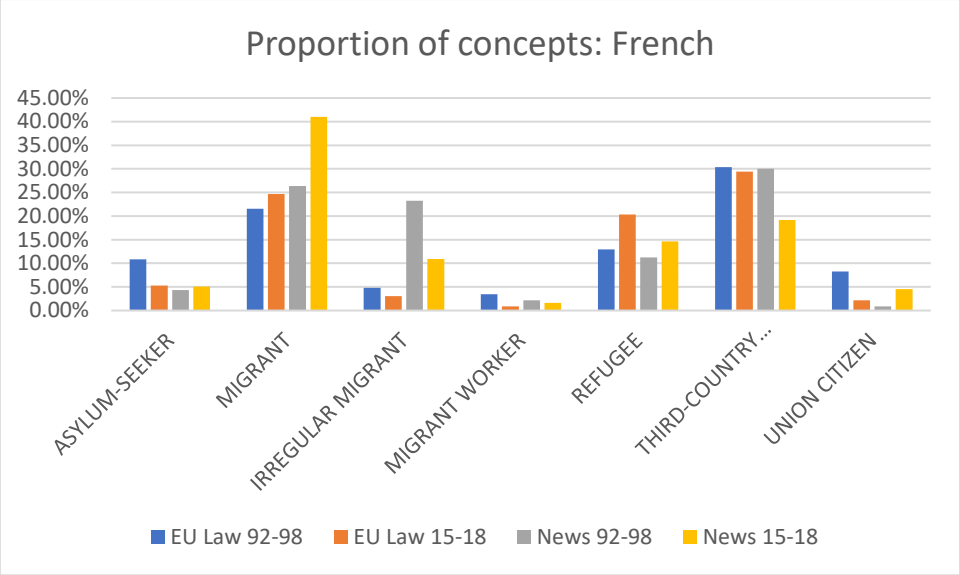


Figure 26: Each concept in each subcorpora as an overall proportion of French terms

In opposition to the findings for English above, there is a closer alignment in the trends for concepts in French across the two text types. Five of the seven concepts, MIGRANT, IRREGULAR MIGRANT, MIGRANT WORKER, REFUGEE and THIRD-COUNTRY NATIONAL, show similar trends in both EU legal texts and newspaper articles, while only ASYLUM-SEEKER and UNION CITIZEN exhibit contrasting trends. Although the extent of increases and decreases in proportional representation of concepts vary across text types (e.g. THIRD-COUNTRY NATIONAL decreases by 0.98 percentage points in EU legal texts compared to 10.84 percentage points in newspaper articles), the fact that the majority of the concepts are moving in the same direction suggests greater similarities in the general subjects covered by EU legal texts and news articles in French.

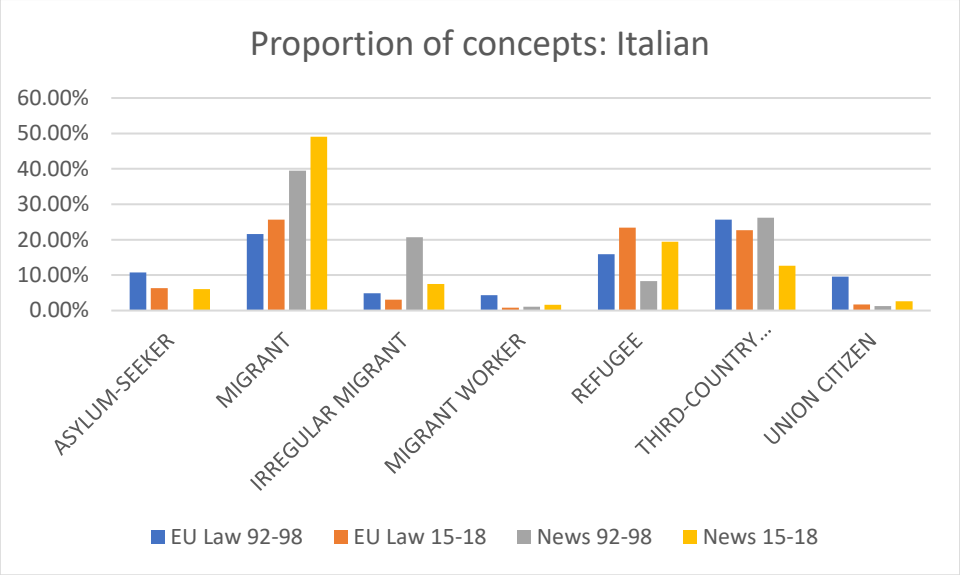


Figure 27: Graph showing each concept in each subcorpora as an overall proportion of Italian terms

With regard to the findings for English and French above, Italian falls in between the two in terms of the similarity in the trends exhibited by the two text types. Across the Italian corpora there are four concepts, MIGRANT, IRREGULAR MIGRANT, REFUGEE and THIRD-COUNTRY NATIONAL, which show the same trend in EU texts and newspaper articles, while the other three, ASYLUM-SEEKER, MIGRANT WORKER and UNION CITIZEN display contrasting trends. This data on how closely aligned the two text types are in each language with regard to the subjects covered and the concepts referenced is a useful initial observation since it indicates where there is divergence at concept level before analysing the terms used within each conceptual category.

6.3 Quantitative results on terminological convergence/divergence

This section applies the information collected on term frequency to establish the outcome of the three convergence/divergence indicators described in section 6.1

above. Figures 28, 29 and 30 below shows the results of these three indicators for each language over both time periods.

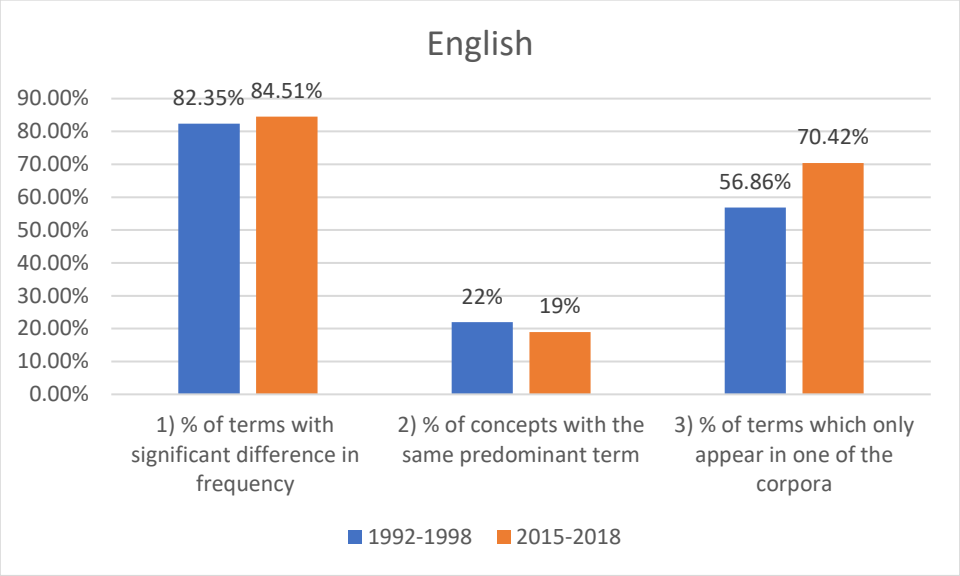


Figure 28: Results of indicators 1, 2 and 3 for English for both time periods

For English, the first indicator showing the percentage of terms with a statistically significant difference in frequency across the two text types shows a statistically significant frequency difference equating to 82.35% of terms, which increases slightly to 84.51% by the later time period. This suggests a minor divergence in the migration terminology used in English in these two spheres. The results of indicator 2 for English corroborate this finding, with a slight decrease in the number of concepts with the same predominant term, falling from 22% to 18%. This indicates terminological divergence over the time period since there is a higher proportion of concepts in the later period where the predominant terms in EU legal texts and newspaper articles differ. Indicator 3 is also in line with the other two, since the percentage of terms that only appear in one of the two corpora rises substantially over the time period. Consequently, for English, these three indicators provide a

uniform indication of terminological divergence over the time period for EU legal texts and press articles in English.

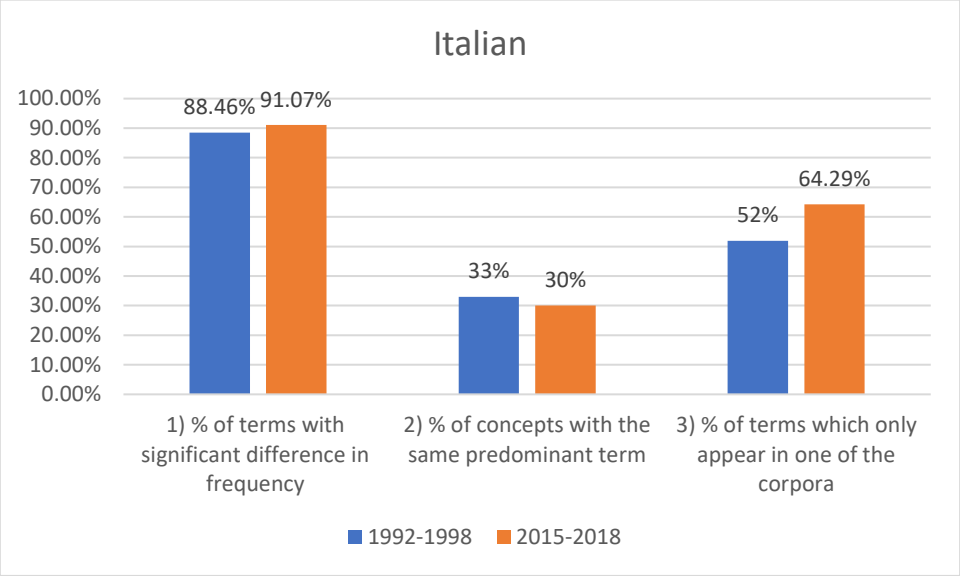


Figure 29: Results of indicators 1, 2 and 3 for Italian for both time periods

As shown in figure 29, Italian shows a very similar pattern to English. With regard to indicator 1, it has the highest degree of terminological difference between EU and newspaper texts in the earlier time period, standing at 88.46% of the overall terms showing a statistically significant difference in frequency. This then increases slightly further over the time period, to 91.07%, indicating further divergence in terminology used in the two text types. Indicator 2 shows a decrease in the percentage of concepts sharing the same predominant term across EU legal texts and press articles in Italian, thus further strengthening the suggestion of terminological divergence across the two text types. The third indicator reinforces this trend, showing a rise in the percentage of terms which only appear in one of the two corpora from 52% to 64.29%. Italian therefore follows a very similar trend to English, with all three indicators pointing towards divergence of migration terminology used in EU legal texts and press articles over the time period.

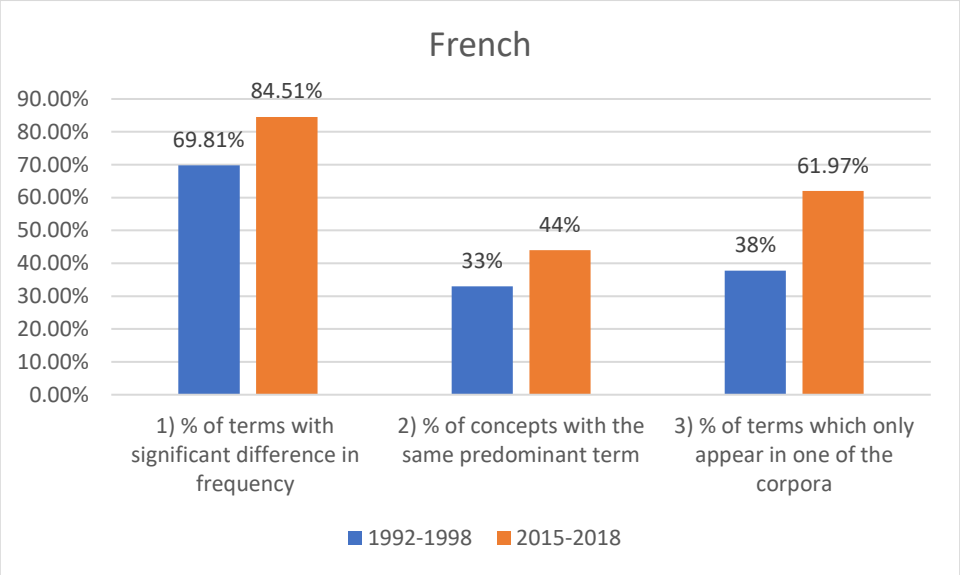


Figure 30: Results of indicators 1, 2 and 3 for French for both time periods

In contrast to the results for English and Italian, the indicators for French show a slightly different and more mixed picture. The results in the earlier time period show a smaller terminological difference between EU legal texts and newspaper articles regarding the migration terminology used compared to the other two languages, with the percentage of terms showing a significant frequency difference standing at 69.81%. However, this percentage rises significantly to 84.51%, becoming identical to the percentage for English in the same time period, suggesting substantial divergence between terms used in the two text types in French over the period. The second indicator is markedly different for French, since it shows a significant increase (from 33% to 44%) in the percentage of concepts which have the same predominant term in both EU legal texts and press articles. This is in stark contrast to English and Italian for this indicator and suggests a level of convergence regarding the terminology used in French. Indicator 3 for French confirms the findings from

indicator 1, suggesting terminological divergence since there is a substantial increase in the percentage of terms which only appear in one of the two corpora, rising from 38% to 61.97%.

Consequently, the results for French are less conclusive than for English and Italian, with indicators 1 and 3 pointing towards divergence, while indicator 2 suggests a level of convergence. It appears upon analysis of the results that this anomalous result for French with indicator 2 can be largely attributed to the fact that a significant number of more specialist migration terms that feature in EU legal texts are also used in French newspaper texts, while their English and Italian equivalents are confined solely to the EU texts. Instances of such terms include '*travailleur détaché*' [posted worker] and '*apatride*' [stateless person], as per the following examples below taken from French newspaper articles in the later period:

> *Je crois que l'Europe qui a été construite, c'est une Europe de la violence sociale, comme nous le voyons dans chaque pays chaque fois qu'arrive un* ***travailleur détaché****, qui vole son pain aux travailleurs qui se trouvent sur place*
>
> [*I think that the Europe which has been created is a Europe of social violence, as we see in every country every time a* ***posted worker*** *arrives and takes the jobs of local workers*.]
>
> (Le Figaro, 2017)
>
> *Pour la seconde, il s'agit, par exemple, de substituer un premier titre de séjour de quatre ans à un titre d'un an pour les étrangers qui bénéficient d'une protection subsidiaire ou qui sont* ***apatrides***,…

> [*For the second, it is a case, for example, of replacing an initial four-year residency permit with a one-year permit for foreigners enjoying subsidiary protection or who are* ***stateless persons***…]
>
> (Le Figaro, 2018)

The inclusion of such terms in newspaper articles therefore contributes to the higher proportionate number of shared predominant terms across both text types in French than in the other two languages and suggests more readiness among French newspapers to use technical migration vocabulary. In addition, the results set out in section 6.2 above also demonstrate a closer relationship between the concepts covered by the two different text types in French compared to the other two languages, which would therefore be another contributing factor to the more complex picture identified for French terminological difference between EU legal texts and press articles.

Taken together, the quantitative results set out above indicate clear terminological divergence in migration terminology between EU legal and newspaper texts for English and Italian and a more complex picture for French. Indeed, the results for French show a higher proportion of concepts where EU legal texts and newspaper articles share the same predominant term in the later time period than in the earlier one. An analysis of the results also shows more terminological similarity in the earlier time period for French EU legal texts and newspaper articles than for English or Italian ones. This is demonstrated by the significantly lower percentage of terms with a statistically significant frequency difference across the text types as well as the significantly lower percentage of terms which only appear in one of the two corpora.

This could be explained, at least in part, by the fact that, although English was already widely used during this earlier period, French was still predominant as the main drafting language of EU legislation from which the other language versions were ultimately derived. Given the language trends within the EU in the 1990s, as discussed in chapter 2, many of the terms used would have gone through a process of primary term creation in French (cf. Fischer, 2010). These French terms are therefore likely to have been more readily comprehensible to general speakers than the English and Italian terms derived from them. An example of such a case taken from the earlier part of the EU corpus is the terminology used in that period for the concept CROSS-BORDER WORKER. In the 'Proposal for a Council Act establishing the Convention on rules for the admission of third-country nationals to the member states' (COM 97(387)), the French term used for this concept is '*travailleur transfrontalier*' while the English version uses '*transfrontier worker*'. The French term here is commensurate with standard French (and remains the accepted term for this concept to this day), while the English term is clearly a calque from the French. It would not necessarily be understood by the general English speaker and would not therefore be used in non-specialist press articles on the matter. It is interesting in this context therefore that a similar effect does not appear to occur when English becomes the predominant language of primary term creation in the EU in the later time period. Indeed, there is no evidence of any convergence between EU and news migration terminology in English according to these general quantitative results, demonstrating the extent to which by this point EU English (along with all other EU language varieties) has developed its own distinct terminology, as discussed in more detail in section 6.5, removed from standard varieties used in the non-specialist press (cf. Mori, 2018a).

The above results are based on overall quantitative findings which combine the data derived from each conceptual category to provide a picture of the general trends. Subsequently, the below sections set out a complementary analysis of the conceptual categories arising from this comparative corpus study to understand in greater depth the main terminological dynamics present across time periods and text types for the three languages involved.

6.4 Terminological trends

6.4.1 THIRD-COUNTRY NATIONAL

An analysis of the terms contained within the category THIRD-COUNTRY NATIONAL shows relatively uniform terminological divergence across the three languages with the terminological trends for this concept in the two text types moving in entirely different directions. In EU texts, there is a clear consolidation of the use of the term '*third-country national*' in English and its French and Italian equivalents '*ressortissants de pays tiers*' and '*cittadini di paesi terzi*', as shown in figure 31 below.

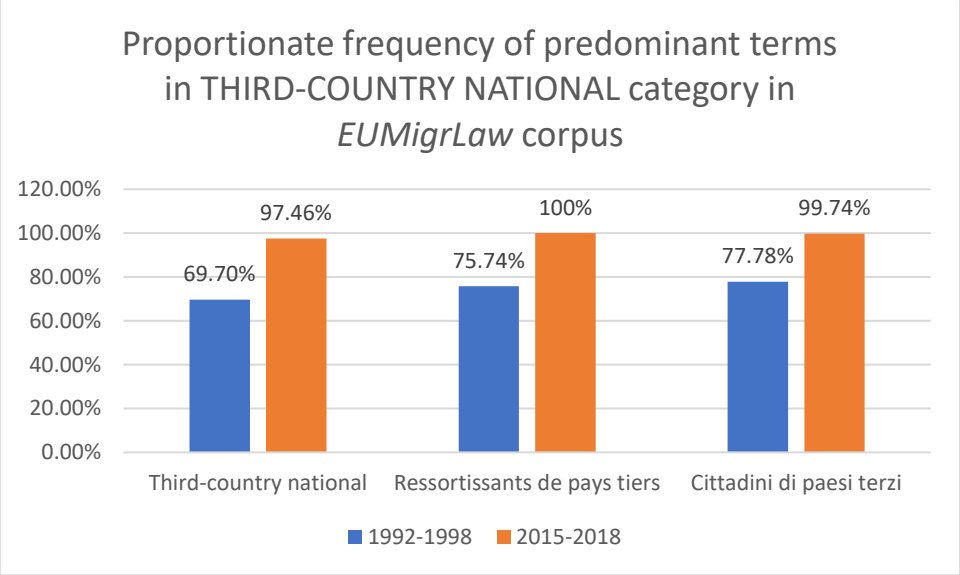


Figure 31: Change in prevalence of predominant term in conceptual category THIRD-COUNTRY NATIONAL in the EUMigrLaw Corpus

All three languages show a clear proportionate increase in the use of this term over the time period, resulting in the almost complete absence of any alternative terms in this category in the later time period. In contrast, this conceptual category for news articles shows both greater terminological instability for English and French indicated by an increase in total terms (5>8 in English and 4>7 in French), and for Italian, increased prevalence of a different predominant term '*straniero*' [foreigner] (rising from 49.67% of the category to 88.28%). The extent of the divergence in terminology between the two text types is highlighted by the reduction in the number of terms which appear in both corpora. In the earlier time period, taking all three languages together, half of all terms (10 of 20) in this category appear in both the EU legal text corpus and the newspaper corpus. However, by the later period, only 3 out of a total of 24 terms appear in both corpora. This indicates a very substantial deviation in terminological choices taken in the two text types across all languages studied.

The significant divergence in this conceptual category across the languages would seem to be a consequence of the markedly different objectives of the two text types. Indeed, the newspaper articles in all three languages in both time periods display a preference to varying degrees for the English term '*foreigner*' and its French and Italian equivalents '*étranger*' and '*straniero*', as can be seen from the below examples, all taken from the later newspaper subcorpus.

> *Ms Rudd claimed she had been "very thoughtful" in wording the speech, in which she suggested* ***foreigners*** *were taking jobs British people could have – despite the UK currently having record levels of employment*
>
> (The Independent, 2016)
>
> *Il est aussi très compliqué, au regard du droit européen, de couper les* ***étrangers*** *installés en France de leurs ascendants directs*
>
> [*It is also very complicated, in light of European law, to cut off* ***foreigners*** *settled in France from their direct ascendants*]
>
> (La Croix, 2015)
>
> *Gli* ***stranieri*** *respinti alla frontiera sono stati 5,161*
>
> [*5,161* ***foreigners*** *were pushed back to the border*]
>
> (Corriere della Sera, 2018)

Writing for non-specialist audiences, journalists aim to report information and communicate ideas without relying too heavily on technical terminology, which may alienate readers and be less emotively effective. Loyalty to terminological accuracy

is not a high priority. Indeed, as pointed out in the previous chapter, some newspapers may even choose to blur terminological boundaries and opt for terms with clearly negative connotations in order to serve their ideological purposes. In addition, as discussed in the previous chapter and shown by the above examples, news articles often do not distinguish between foreign nationals from other EU member states and those from outside the EU. This distinction is often not of central importance to the news articles in the corpus and the EU's official term, '*third-country national*', is likely to be deemed as highly specialised jargon which would not be understood by large numbers of readers, hence why even more progressive media outlets have not chosen to embrace this terminology in more recent times.

6.4.2 IRREGULAR MIGRANT

Another concept which demonstrates significant terminological divergence between the two text types over the time period is IRREGULAR MIGRANT. Among the most frequent terms in this category for English, the percentage difference in the relative frequency of '*illegal immigrant*' between the two text types rose from 160.88% to 185.40% and '*illegal migrant*' increased from 100.99% to 190.40%. Equally, the term '*irregular migrant*', records a percentage difference between the text types of 183.43% by the later time period. In addition to these growing percentage differences, the changes in the proportions of terms within the category also indicate divergence. Indeed, '*illegal immigrant*' appears as the predominant term in the earlier time period in both corpora, representing 74.36% of the category for EU texts and 60.08% for newspaper texts. However, over the time period, while this term has consolidated its predominant position in newspaper texts, increasing to a proportion

of 73.24%, the predominant term in the EU texts has shifted entirely to '*irregular migrant*', which represents 81.03% of the overall number of terms in the category. These results indicate that both the relative frequencies of the terms and the proportions that the terms represent within the category have diverged significantly over the time period, suggesting markedly different terminological directions being taken by those producing the two different text types. Indeed, as shown in chapter 3, in the EU, there has been a very significant move away from the use of the adjective '*illegal*' in this context, towards the alternative '*irregular*'. Such a wholesale change however has not been forthcoming in newspaper articles, despite some small signs of a shift from certain outlets. Nevertheless, despite these strong indicators of divergence in this category for English, the gradual introduction of terms including the adjective '*irregular*' in newspaper texts, namely '*irregular migrant*' and '*irregular immigrant*', could point to a small-scale impact of EU terminology on journalists writing on migration, as discussed further in section 6.5 below.

Similar changes are also evident in French in the conceptual category IRREGULAR MIGRANT. In French, 71% (10 of 14) of the overall terms in the category appear in both corpora in the earlier time period. By the later time period, this percentage drops to just 29% (5 of 17). Therefore, by the later time period, there is a majority of terms which belong either only in the EU corpus or only in the newspaper corpus, providing a strong indication of divergence where many of the common terms in one text type simply do not appear at all in the other. There is also no evidence of any convergence in the predominant terms, with the adjective '*irrégulier*' growing in prevalence in the EU texts, while '*clandestin*' remains predominant in the newspaper texts. A similar pattern with the adjectives '*irregolare*' and '*clandestino*' is also apparent for Italian in this category.

Nonetheless, much like for English above, the increase in the use of '*irrégulier*' and '*irregolare*' in the newspaper corpus for French and Italian in the later time period, although not predominant, may indicate a level of influence from EU terminology. The below examples show how these terms are used in the later time period in the newspaper corpus.

> *I migranti **irregolari** sono i più vulnerabili*
>
> [***Irregular** migrants are the most vulnerable*]
>
> (Corriere della Sera, 2018)
>
> *Depuis la loi Chevènement Jospin du 12 mai 1998, la France offre un dispositif particulièrement généreux aux étrangers en situation **irrégulière** sur le territoire français atteints d'une maladie grave*
>
> *Since the Chevèment Jospin Act of 12 May 1998, France has offered particularly generous measures to foreigners in an **irregular** situation on French territory if they are suffering from a serious illness*
>
> (Le Figaro, 2015)
>
> *The academic study found little evidence that immigration enforcement activity deters **irregular** migrants or encourages them to leave the country*
>
> (The Independent, 2018)

Figure 32 below illustrates how the adjectives '*irregular*', '*irrégulier*' and '*irregolare*' have grown in prevalence in both corpora over the time period.

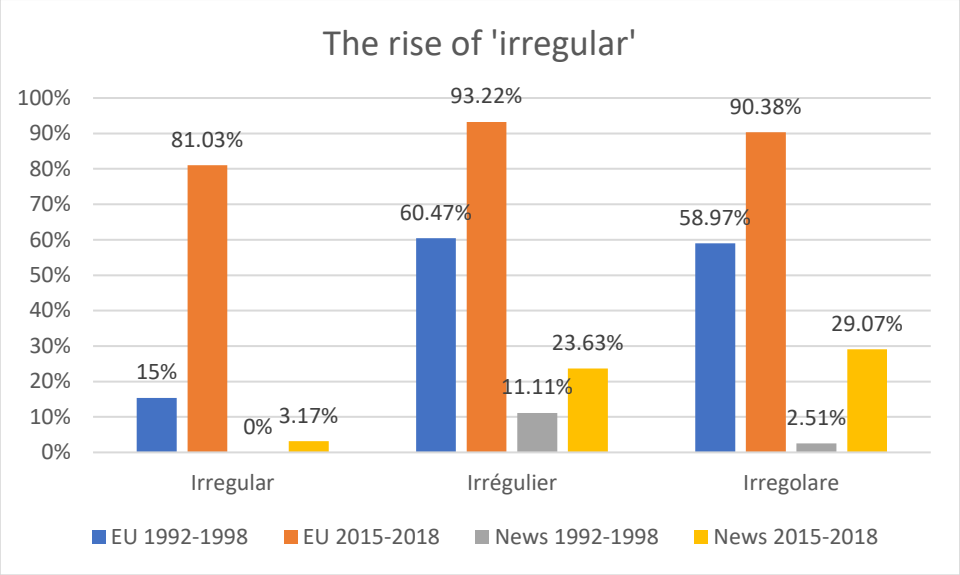


Figure 32: Proportion of terms in conceptual category IRREGULAR MIGRANT containing adjectives 'irregular', 'irrégulier' and 'irregolare' in EU texts and news articles

More specifically, the EU texts appear to have been forerunners in using this terminology in the earlier time period, when such terminology remained uncommon in newspaper texts across the languages. It is therefore possible that this usage in EU legal texts paved the way for increased usage in newspaper texts in the later time period, which saw an admittedly modest increase in English, but much more significant rises in French and Italian. Consequently, although the key indicators for this concept point to overall divergence in terminology used in each language across the time period, with regard to the specific usage of the adjective '*irregular*' in this category, there does appear to have been a level of influence, albeit minor, on the terminology used by certain newspapers.

6.4.3 UNION CITIZEN

A third concept that exhibits a similar trend of general terminological divergence as seen with THIRD-COUNTRY NATIONAL and IRREGULAR MIGRANT is UNION

CITIZEN. In French, the predominant term in newspapers, '*Européen*', becomes increasingly prevalent over time in this category, but shows a negligent increase in EU texts, while the predominant term in the EU corpus, '*citoyen de l'Union*', fails to appear in the newspaper corpus at any point. The results for French therefore demonstrate how the two corpora are moving in different directions with regard to terminological preference for this concept. A similar pattern can be observed for Italian, where the predominant term in the newspaper corpus, '*Europeo*', is completely absent from the EU corpus in the later time period, while the favoured term in EU texts, '*cittadino dell'Unione*', also fails to appear at all in the newspaper corpus in the later time period. It would appear that the more general term '*Européen*'/'*Europeo*' is increasingly preferred among all sections of the French and Italian press to express the concept of UNION CITIZEN moving away from any explicit reference to the concept of citizenship, thus widening the gap in the frequency of the most prevalent terms in this category for both French and Italian.

Nevertheless, English results for this concept are more ambiguous. As for the other languages, the percentage difference between the frequencies of the most commonly used terms also grows over the time period. Indeed, in the earlier time period, the frequency difference of term '*EU citizen*' between EU legal texts and press articles (67.55%) cannot be deemed statistically significant according to the log likelihood values, but this percentage increases substantially (166.59%) by the later time period. This is also the case for the term '*European*', whose percentage difference between the two text types rises from 159.54% to 174.13% by the later time period. However, as shown in Table 21 below, it is important to note that for English, by the later time period, the two corpora share the same predominant term: '*EU citizen*'.

Table 21: Predominant terms in English for UNION CITIZEN conceptual category

	EUMigrLaw Corpus	***MigrationInTheNews Corpus***
1992-1998	Citizen of the Union (44.12%)	European (54.29%)
2015-2018	EU citizen (28.57%)	EU citizen (68.45%)

This would suggest that, despite the very substantial frequency differences between terms, there is in fact a level of convergence in this category for English since both corpora have the same preferred term for this concept. The below examples show how this predominant term is used in both corpora in the later time period.

> *The proposed amendments will make systematic checks of **EU citizens** against databases at all external border (air, sea and land) – which are today possible based on a risk assessment – compulsory*
>
> (Communication from the Commission to the European Parliament and the Council, 2015.)

> *The prime minister's intervention was a sign of her determination to end free movement of **EU citizens** after Brexit but will dismay business leaders*
>
> (Financial Times, 2018)

This also demonstrates the importance of combining the above measures of divergence and convergence in order to gain a full picture of the dynamics within and between the corpora. Indeed, although by most measures the terminology for this concept has diverged over the time period, there is some evidence for English that while there is a widening gap in terms of relative frequency of terms, the proportion

of those terms within the category can in fact demonstrate some level of convergence.

6.4.4 ASYLUM-SEEKER and REFUGEE

Despite the general quantitative results showing overall divergence in terminology between text types across all three languages, there are a number of concepts that display significant levels of convergence over the period. Two of those categories are ASYLUM SEEKER and REFUGEE. In the English parts of the corpora, the terms '*asylum-seeker*' and '*asylum seeker*'[16] both display a statistically significant difference in frequency across the text types in the earlier period, but the frequencies of these terms in the later time period grow similar enough for them to no longer be deemed statistically significant according to their log likelihood values, as shown in figure 33 below. Moreover, there is also convergence regarding the predominant term over the time period. In the earlier period, the EU texts favour '*asylum-seeker*' compared to the newspapers mainly using '*asylum seeker*'. By the later time period, both corpora have a clear preference for the term '*asylum seeker*'. For the conceptual category REFUGEE, the percentage difference between the frequency of the predominant English term '*refugee*' across the text types also decreases significantly over the time period, falling from 128.62% to 40.56%. All of these indications point to a convergence in the terminology being used to express these two concepts across the two text types, both in the overall frequency with which they are used and proportionately as part of the category in question.

[16] As discussed in chapter 3, although there is no semantic difference in cases such as these, orthographic differences are counted as different terms for the purposes of this study.

This same observation for these concepts can also be made for Italian. The percentage difference between the relative frequency in EU legal texts and press articles of the predominant term in the category ASYLUM SEEKER, '*richiedente asilo*', decreases from 191.26% in the earlier time period to 57.49%, as shown in figure 33. Similarly for the category REFUGEE, the percentage difference in frequency of the term '*rifugiato*' across the text types falls from 52.50% to 31.18%. In addition, the term '*rifugiato*' becomes the predominant term in both corpora by the later time period, representing 90.48% of the category in the EU corpus and 54.26% in the newspaper corpus. In so doing, this term overtakes '*profugo*' in both cases, thus reflecting a common trend for Italian shared across both text types regarding the terminology used to express this concept. This shift is brought into particularly sharp focus in the EU corpus by the different ways in which the 1951 Refugee Convention is referenced in the two time periods, as per the below examples.

> …*considerando che un simile approccio equilibrato deve essere basato sulle disposizioni vigenti nel settore, con particolare riferimento alla Convenzione di Ginevra sullo stato dei* ***profughi***[17] *del 1951*…
>
> […*given that a similar balanced approach must be based on the provisions in force in the domain, with particular reference to the 1951 Geneva Convention on the status of* ***profughi***…]
>
> (Resolution on the Communication from the Commission to the Council and the European Parliament on immigration and asylum policies, 1995)

[17] As discussed in section 5.2.1, *profugo* is a more general term than *rifugiato*, since it refers to a person who has fled their home country for any reason whatsoever, yet does not necessarily fulfil the criteria to obtain refugee status.

> *L'applicazione della presente direttiva non pregiudica gli obblighi derivanti dalla convenzione di Ginevra, del 28 luglio 1951, relativa allo status dei* ***rifugiati****…*
>
> [*The application of this directive is without prejudice to the obligations deriving from the Geneva Convention of 28 July 1951 regarding the status of* ***refugees****…*]
>
> (Proposal for a Directive of the European Parliament and of the Council, 2018)

As for English and Italian, the French terms for the concept REFUGEE also display convergence, with the percentage difference in the frequency of the predominant term, '*réfugié*', between EU legal texts and press articles falling from 74.73% to 19.41%, as is also shown in figure 33 below.

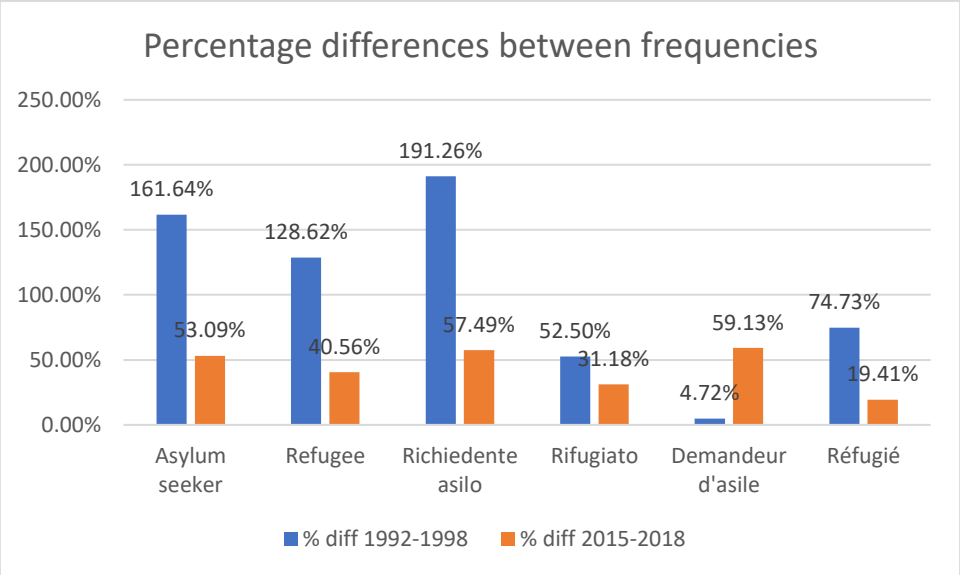


Figure 33: Percentage differences between frequencies of terms across the EUMigrLaw Corpus and the MigrationInTheNews Corpus for both time periods for concepts ASYLUM-SEEKER and REFUGEE

The one exception to this trend across these two concepts is the French term '*demandeur d'asile*' in the category ASYLUM SEEKER, which records a significant increase in the percentage difference of frequencies over the time period, as shown in figure 33. Indeed, in the earlier time period this value stands at just 4.72% but rises to 59.13% by the later period. This is due both to a fall in the frequency of this term in the EU corpus and a rise in the newspaper corpus, combined with the introduction of alternative terms in the EU corpus, '*demandeur de protection internationale*' and '*demandeur d'une protection internationale*', which constitute 16.50% of the overall number of terms in the later subcorpus. Despite this single outlier, the results for these two concepts indicate consistent decreases in the percentage differences between frequencies of terms across the two text types for the key terms in all three languages in the conceptual categories ASYLUM SEEKER and REFUGEE, suggesting a clear terminological convergence in these particular categories over time.

6.4.5 MIGRANT and MIGRANT WORKER

The results for the conceptual categories MIGRANT and MIGRANT WORKER show significantly different trends for each of the three languages. An analysis of the terminology for French indicates a level of convergence for both concepts. Indeed, the French parts of the two corpora have different predominant terms in the earlier time period for the concept MIGRANT, namely '*immigrant*' in the EU corpus (33.78%) and '*immigré*' in the newspaper corpus (47.08%). By the later period, the predominant term '*migrant*' is the same in both, representing 94.82% of the category

for the EU corpus and 53.18% of the newspaper corpus, as shown in Table 22 below.

Table 22: Predominant terms in French for conceptual categories MIGRANT and MIGRANT WORKER

	EUMigrLaw Corpus	***MigrationInTheNews Corpus***
1992-1998	*Immigrant* (33.78%) *Travailleur migrant* [Migrant worker] (70.97%)	*Immigré* (47.08%) *Travailleur immigré* [Immigrant worker] (38.78%)
2015-2018	*Migrant* (94.82%) *Migrant économique* [economic migrant] (41.18%)	*Migrant* (53.18%) *Migrant économique* [economic migrant] (25.00%)

A similar trend is also evident for French with the category MIGRANT WORKER, also shown in Table 22 above. The two text types have different predominant terms in the earlier time period, with '*travailleur migrant*' (70.97%) the preferred term in the EU corpus while '*travailleur immigré*' (38.78%) is favoured in the newspaper corpus. By the later time period, however, the predominant term in both corpora for this concept in French becomes '*migrant économique*', representing 41.18% of the category in the EU corpus and 25.00% in the newspaper corpus, as per the below examples taken from the later time period in both corpora.

> *D'autre part, cela implique de faire la distinction entre réfugiés politiques, éligibles à l'asile, et* ***migrants économiques****, qui ne le sont pas*
>
> [*Moreover, that involves distinguishing between political refugees, who are eligible for asylum, and* ***economic migrants****, who are not*.]
>
> (Le Monde, 2017)

> *Étant donné l'importance de l'afflux de migrants et pour respecter la convention de Genève et les principes généraux des droits de l'homme, il importe de distinguer clairement les* ***migrants économiques*** *des réfugiés*
>
> [*Given the scale of the inflow of migrants and in order to comply with the Geneva convention and the general principles of human rights, it is important to clearly distinguish between* ***economic migrants*** *and refugees*.]
>
> (Opinion of the European Economic and Social Committee, 2016)

In addition, the number of terms displaying a statistically significant difference in frequency between EU legal texts and press articles in this category for French falls over the time period. In the earlier time period, 4 of the 5 terms showed a statistically significant frequency difference based on the log likelihood values, but this falls to just 3 of 7 by the later time period. Taken together, these various indicators demonstrate terminological convergence for French for these two conceptual categories.

There is a more mixed picture for Italian regarding the terminological trends in these two categories. The category MIGRANT WORKER produces a similar result as for French. The predominant terms differ in the earlier time period, with '*lavoratore migrante*' [migrant worker] being preferred in the EU corpus and '*lavoratore straniero*' [foreign worker] being favoured in the newspaper corpus. By the later period, there is a convergence over the preferred term since '*migrante economico*' [economic migrant] becomes the most prevalent term in both corpora. Nonetheless, the convergence of Italian terminology in this category is not reflected for the concept MIGRANT, which presents a complex series of dynamics. The predominant terms

are the same in both corpora but change over the time period, meaning that both text types prefer the term '*immigrato*' [immigrant] in the earlier time period and '*migrante*' [migrant] in the later one. This indicates a similar shift occurring in both corpora but does not necessarily indicate convergence. With regard to frequencies of terms, although the percentage difference between the frequencies of the term '*migrante*' in the two text types falls from 106.47% to 51.64% the same metric for the term '*immigrato*' rises from 153.28% to 188.17%. The indicators for Italian therefore show a complex dynamic for Italian, with some indication of convergence for the concept MIGRANT WORKER but inconclusive results regarding convergence or divergence for the concept MIGRANT.

Also distinct from French and Italian, the English results for the concepts MIGRANT and MIGRANT WORKER indicate overall terminological divergence in both. Indeed, while both the legal and news corpora have the same predominant term in the earlier time period, '*immigrant*', they diverge in the later one, with the EU corpus preferring '*migrant*' while the newspaper corpus retains '*immigrant*'. Despite a slight fall in the percentage difference in the frequency of the term '*migrant*' between the two text types over the period (56.34% to 39.98%), the difference in the frequency of the term '*immigrant*' increases to a greater extent (134.17% to 190.16%). Likewise for the concept MIGRANT WORKER, the difference in frequency of some of the most prevalent terms in this category increases, including '*migrant worker*', which rises from 118.71% to 159.19% and foreign worker, increasing from 28.78% (and deemed not statistically significant) to 163.29%. The predominant term in the early time period for this concept in the EU corpus is '*migrant worker*' (70.97%), which then shifts to '*economic migrant*' (53.33%) by the later time period. Meanwhile, the trend for the newspaper corpus moves in the opposite direction, with the term '*economic*

migrant' (59.09%) being preferred in the earlier time period while '*foreign worker*' (37.14%) becomes predominant in the later one. Although there is a significant decrease in the frequency difference of the term '*economic migrant*' over the time period (from 156.05% to 11.84%), the overwhelming majority of indicators for English in these two categories suggest that the terminology for the concepts MIGRANT and MIGRANT WORKER has diverged between EU legal texts and newspaper articles over the time period in question.

6.5 Discussion

The general quantitative results for the comparative analysis of migration terminology used in EU legal texts and press articles in English, French and Italian presented here indicate overall divergence in the terms used in the two text types over the time period. Examining each language individually, English and Italian display a consistent divergence according to all the indicators used, while French shows a slightly more ambiguous picture with one of the three indicators pointing to a level of convergence. This is largely due to the higher number of technical terms used in French press articles on the subject of migration compared to English and Italian print media. In any case, the overall trend of terminological divergence that emerges from this corpus analysis indicates therefore a widening of the gap between the migration terminology used in EU legal texts and in press articles over the time period being studied. This in turn indicates that the EU migration terminology explored in chapter 4 has had no substantial impact on the terminology used in non-specialist contexts, such as the print media. By almost every metric, the terms used to describe most key concepts in the field of migration in these two text types have

grown more different. Reflecting on the initial research aims, these results would suggest that, although there is evidence, as argued in chapter 4, of translation-induced language change influencing terminological development in EU legal texts, this has generally not crossed over to impact the terminology used outside that sphere. Indeed, these results also therefore indicate that the terminological changes in the English, French and Italian print media, highlighted in chapter 5, cannot in general be attributed to influence from EU legal texts.

Previous research into translation-induced language change has generally focused on the existence and effects of this phenomenon within a single genre. This research is distinct in also attempting to detect the impacts of translation-induced language change beyond the genre in which such change initially occurs. Such a phenomenon would involve terms used in specialist texts transferring to non-specialist language communities through a process of de-terminologization (Meyer & Mackintosh, 2000). When such a transfer occurs, either the term retains its original fundamental meaning or it undergoes a dilution of its original meaning to allow for a layperson's shallower understanding (ibid.). As discussed in chapter 2, there are a number of factors that may encourage such a carry-over to occur. The fact that many press articles on the topic of migration directly or indirectly address EU measures and regulations on the matter could lead to the terms used in those regulations being coopted by journalists writing on the subject. More generally, the prestige of EU law as an authoritative source could also play a role in terms being transmitted beyond the boundaries of the legal texts in which they are used.

Indeed, notwithstanding the general quantitative observations set out above, closer inspection of the individual concepts appears to show that there are some specific, isolated areas of terminological convergence between the text types in each

language. While it is important to repeat that changes in language can very rarely be attributed to a single causal factor, the fact that certain migration concepts in press articles have seen a shift in their terminological profile moving significantly closer to terminology used in EU texts over the time period suggests a level of influence being exerted by those texts and the terminology contained within them. The increased use in later press articles of the adjectives '*irregular*', '*irrégulier*' and '*irregolare*' in the conceptual category IRREGULAR MIGRANT comes in the wake of the already heightened use of these adjectives in EU texts, indicating a gradual reframing of the terminology around this concept starting in EU legal texts and carrying over to less specialist texts. There are also instances within each language of terminological convergence, namely in the categories MIGRANT and MIGRANT WORKER for French, and REFUGEE and ASYLUM SEEKER for English and Italian. For example, the Italian term '*richiedente asilo*' [asylum seeker] appears to have undergone de-terminologization in newspaper texts over the time period. Although the term is common in EU texts in both periods, it is almost entirely absent from the earlier period in the Italian part of the news corpus. By the later period, however, it appears with much greater frequency, as per the below example:

> *Ciò consente di coinvolgere **i richiedenti asilo** in progetti di formazione che i migranti affrontano proprio con l'obiettivo arrivare ad un inserimento lavorativo*
>
> [*This allows **asylum seekers** to get involved in training that migrants have to negotiate in order to find a way into the jobs market*].
>
> (Corriere della sera, 2018)

In most cases, this term appears to have largely retained its original fundamental meaning when used in press articles - a person awaiting the outcome of their

application for protection – and is often contrasted against more general terms, such as '*migranti*', as in the above example.

Another instance of this phenomenon of de-terminologization can be seen with the appearance in newspapers of the English term '*irregular migrants*', as per the below example:

> *The academic study found little evidence that immigration enforcement activity deters* ***irregular migrants*** *or encourages them to leave the country*
>
> (Independent, 2018)

Although not the predominant term for this concept in the later part of the newspaper corpus, it has nonetheless entered the English terminological profile (along with '*irregular immigrant*'), reflecting the extensive use of '*irregular*' in the EU corpus. The use of '*irregular*' in this context, previously confined to EU legal texts and resulting in part from contact with French as described in chapter 4, has therefore begun to be de-terminologized and is becoming increasingly comprehensible to a general lay readership. These findings therefore indicate that it is possible, in certain specific circumstances, for terminology that has been the subject of translation-induced language change at a specialised level to cross genre boundaries and ultimately spread into non-specialist, general speech communities. In this way, for certain isolated cases, it is possible to establish a potential traceable lineage of influence from EU legal translation to non-translated newspaper texts.

However, the findings indicate that in most cases for migration terminology in these two particular genres (EU legal texts and newspaper articles) and across this particular time periods (1992-1998 & 2015-2018) translation-induced language change does not carry over from the specialist genre in which the changes originate

to a more general genre. Indeed, it appears that opposing factors carry greater weight in this context, preventing or obstructing migration terms in EU texts from being adopted in press articles. On a general level, there is evidently significant distance between the linguistic norms of legal texts and journalistic press articles. Indeed, legal texts are characterised by highly specialized terminology, complex structures and often opaque language, which differs greatly from the style of press articles, often broken down into ‘hard news’ containing the key factual information on an event and ‘features’ that provide more background and provide opinions through editorialization (Bell, 1991, pp. 14). News articles tend to be characterised by a generalisation, clarification and simplification of factual information in order to reduce the amount of material and enhance its newsworthiness (ibid., pp. 65). The objective of journalists is to deliver a product their readers will want to consume, basing their decisions on what they deem more or less newsworthy for their particular audience (Auel, Eisele & Kinski, 2018, pp. 631). Indeed, the two text types have very different targets and readerships, which requires a very different approach to lexical choices. It appears that these factors have blocked the adoption of terminology used in EU legal texts in most cases.

More specifically on the subject of migration, journalists are likely to have a specific stance or ideology that informs their opinion on the subject and therefore shapes the language choices they make in writing the article. This leads to them eschewing terminological rigour for increased comprehensibility and potentially more connotatively charged terms rather than neutral ones. The lines between what is ‘fact’ and what is ‘comment’ are often blurred. Depending on the stance and type of newspaper, the opinions expressed and language used about the same factual information can vary greatly. It is common for the choice of words in such articles to

generally or contextually express a value judgement (Van Dijk, 1998, pp. 31). These factors therefore appear to play a critical role in preventing the widespread transfer of terms from EU texts into press articles.

Consequently, only those terms which are able to satisfy the particular requirements of the news genre are likely to transfer. As observed in this study, the small number of terms which show convergence across the two text types meet certain specific needs of news articles. Indeed, the 'notion of need' is itself a key accelerating factor in driving language change (Aitchison, 2013, pp. 153-4). For example, among the specific cases observed here, there is a clear 'need' among certain news outlets to find a more neutral and less demonising way of discussing those people who had previously been widely referred to as '*illegal immigrants*', hence the uptake of the terms '*irregular migrant*' and '*irregular immigrant*' and their French and Italian equivalents, following the earlier appearance of the adjective '*irregular*' in EU texts.

The continued widening of the terminological gap identified by these findings is further ammunition to the much-repeated claim that the EU institutions are too distant and opaque, isolated from the citizens which they aim to represent (Follesdal & Hix, 2006). The ongoing evolution of EU language varieties is leading to the development of a body of terminology specific to those varieties and unfamiliar to standard varieties of those languages (cf. Mori, 2018a). For example, the appearance in the later time period of this corpus analysis of terms such as '*applicant for international protection*' and '*beneficiary of international protection*' (in the categories ASYLUM SEEKER and REFUGEE) are demonstrations of the continued development of nuanced EU-specific terminology that is absent from non-specialist press articles. The sense of remoteness of the European Union from its

citizens is a central criticism built around the argument that there is a crucial democratic deficit in the way that the institution operates. As it makes clear in its Joint Practical Guide, the EU itself is very aware that ‘texts peppered with borrowed words, literal translations or jargon which is hard to understand are the source of much of the criticism of Union law, and result in it being regarded as alien’ (European Union, 2015, pp. 19). The results from this study do nothing to allay these concerns. Research shows increasing anti-EU sentiment across Europe, partly due to the feeling that the EU is out-of-touch with the needs of its citizens (Dijkstra, Poelman & Rodriguez-Pose, 2020). Indeed, one of the key drivers of the UK referendum result to leave the European Union was a sense that the institutions failed to reflect those it claimed to represent, which can only be further compounded if the language used by the institutions is deemed ‘foreign’ or alienating, made up as it is of a preponderance of ‘Europeanisms’ specific to EU texts (Mori, 2018a, pp. 20).

Despite this familiar criticism, which the results of this study tend to support, it is important to acknowledge the intractability of the challenge faced by the drafters of EU law with regard to this specific point. Indeed, as discussed in chapter 2, EU lawmakers need to ensure that the terminology they use is not too closely associated with any particular member state, making it almost inevitable that much of the language and terminology used in EU law will have a ‘non-native’ quality in any of the official languages. As such, EU drafters and translators are caught in a bind – they are criticised for using terminology which is deemed opaque and remote, and yet this remoteness is to a large extent a necessity of the political circumstances in which they are operating. This situation is further compounded by the fact that justice and home affairs is an area in which Member States are particularly reluctant to cede

sovereignty, leading to the remoteness of the terminology becoming even more exaggerated.

Moreover, results from chapter 5 show that it is not true that terminology in legal texts is diverging from stable terminology elsewhere. On the contrary, the terminological divergence between the two text types studied here is also attributable, in roughly equal measure, to the changes that have been shown to take place in press articles on migration over time. The widening terminological gap identified by this research is therefore the result of terminological development in both text types moving in different directions. Indeed, it would seem more accurate to characterise this observed phenomenon of terminological divergence as largely the result of natural and inevitable language change due to a combination of social and linguistic factors, which differ significantly across the two genres as a result of the divergent requirements of the two text types (e.g. Aitchison, 2013).

Given the general divergence in terminology between the two text types identified in this part of the study, there would seem to be little evidence of public discourse in the media shaping legal language in this field. In fact, such is the level of divergence over the time period here, notwithstanding the isolated examples of convergence already discussed above, that it could be hypothesised that the language used in the print media can even have the effect of motivating alternative terminology to be used in legal language. For example, terms such as ‘*immigrant*’ or ‘*extracomunitario*’, which acquire negative semantic prosody due to their frequent negative associations in the print media appear to be intentionally avoided in later EU legal texts. Indeed, legal discourse may seek to adopt more neutral terminology in such cases, hence the increased divergence detected here, with media terminology having a kind of

repellent influence on legal language, creating a greater distance between more neutral terms in law and terms with increasingly negative connotations in the press.

In light of the results of the three stages of analysis undergone as part of this study and set out in the previous three chapters, the following chapter reflects on these findings, considers them against the backdrop of the original research questions and draws some overarching conclusions for this research.

7. Conclusion

This thesis has uses a range of methodologies to examine the role of translation in migration terminology in legal texts and press articles in English, French and Italian. It harnessed corpus techniques in a stepwise approach across three stages to produce quantitative metrics and qualitative analyses to provide robust responses to the series of research questions set out at the beginning of this thesis. The first stage examined terminological dynamics in the *EUMigrLaw Corpus* in the three languages across the time periods 1992-1998 and 2015-2018. It identified extremely widespread intralingual and interlingual terminological variation in EU legal texts on the subject of migration. This variation reduces somewhat over the time period, but remains significant nonetheless. The results from this part of the analysis allow us to confirm the first research hypothesis that terminological variation of various kinds can be detected in EU migration law.

The results of the first stage also indicate terminological convergence among the three languages in question, demonstrated by an analysis of the IVI values for each language pair. Indeed, this stage provides evidence of the influence of English on Italian and French terminology in EU legal texts over the time period, shown through lexical, morphosyntactic and connotational changes in the terms used to express a range of concepts in those languages over the period. Consequently, it offers strong evidence of translation playing a role in propagating language change in the case of at least some key migration concepts in EU law. This therefore contributes to answering the second research question by demonstrating that the variation detected can in part be attributed to translation-induced language change.

Stage two examined synchronic and diachronic terminological variation within each language in the *MigrationInTheNews Corpus*. The results of this stage indicated large-scale change within each language regarding the key migration concepts and terms used in UK, French and Italian press articles. These findings provided data to contribute to answering the final research question on the relationship between terminology in the two corpora, pointing to extremely complex and varied dynamics across concepts and languages in the news corpus, with English showing the most substantial change, followed by Italian then French.

The final stage compared the terminological dynamics at play in the two corpora. The goal of this stage was to ascertain the extent of similarity between the changes observed in the previous stages in the form of observable shared patterns which could point to an influence of the changes in EU terminology on news articles. As shown in chapter 6, this stage found widespread divergence in the terminology used for key concepts in migration between EU legal texts and news articles over the time period. The results of this final stage enabled the final hypothesis to be tested, indicating that the translation-induced changes identified in stage one have had little impact on news terminology. Nonetheless, this chapter also identifies a small number of isolated cases where translation-induced changes do appear to have crossed the genre boundary and had an impact beyond EU law. This appears to occur when there is a specific linguistic need which terms in EU law are able to meet.

In light of these findings, this research study contributes to furthering knowledge in the academic fields of Translation Studies, Linguistics, Law and Migration Studies. Building on previous research on the intersection of Translation Studies and Linguistics examining the impact of translation on language change, this study

shows how translation-induced language change plays a role in the terminological dynamics driving the evolution of migration terminology in the languages studied here. This research addresses this phenomenon in a number of original ways. Firstly, rather than examining a language pair as previous studies have done, it explores impacts on three different languages, enabling insights to be drawn on the different ways in which the languages investigated here behave under the influence of translation. This approach also allows for a better conceptualisation of the EU translation process as a complex network of influences, where various different languages can ultimately contribute to shaping terminology in legal documents produced by its institutions.

Furthermore, while research into this phenomenon has examined how it manifests itself in a wide range of different genres, including popular science (Malamatidou, 2011), business (Bisiada, 2013) and news (McLaughlin, 2013) among others, this study instead has a thematic rather than a genre-based focus. Investigating terminology in migration is particularly significant for the real-life impact that the use of language in legal and social contexts can have on the experience of migrants, and this study is original in exploring the role of translation in shaping this terminology in both legal and journalistic texts. Indeed, it explores the potential of translation-induced language change to originate in terminology in one genre and subsequently undergo de-terminologization in order to cross genre boundaries from specialist to non-specialist contexts. While the majority of the findings here indicate that forces work overwhelmingly against such a leap occurring, the few isolated instances that suggest a direct lineage of influence in particular circumstances nonetheless indicate the genre-crossing potential of translation-induced language change. These findings pave the way for future research further exploring this potential, which could

encompass different languages, genres and themes. In general, therefore, this research shows how translation in an institutional setting constitutes a significant factor in a complex network of influences shaping legal migration terminology across different languages, and that this influence is not solely limited to the EU institutional setting but can also be detected further afield.

In the field of Law, this research provides new insights into the evolution of legal terminology in migration-related legislation in the European Union. This study shows how the theory of translation-induced language change can be applied to the EU environment and sets out the implications of this on some of the key legal terms used in this area of law. It demonstrates how the terminology for central concepts in the field evolves, highlighting the somewhat flexible nature of seemingly fixed legal terms in light of a changing legal and policy context. This research builds on other studies that have highlighted the extent of terminological variation in EU legislation (e.g. Prieto-Ramos & Morales-Moreno, 2019; Peruzzo, 2013) by providing a combination of quantitative and qualitative data to demonstrate the degree of terminological instability present in the various language versions of EU legislation. In so doing, it draws attention to the importance of translation and translators in the legislative process. Indeed, as demonstrated in this thesis, despite the extent of terminological resources available in the institutions, the impact of language contact often appears to overcome stringent quality processes to exert influence on terminological decisions and long-term terminological evolution.

Also with regard to Law, this study highlights various instances where legal discourse in all three languages has tended towards more neutral terminology in order to respond to increasingly negative perceptions created by terminology used in press articles. Indeed, divergence in terminology for concepts such as ‘*irregular*

migrant' between the two text types appear in part to be motivated by a willingness in legal circles to avoid terminology that has developed a negative semantic prosody in media discourse.

In light of previous arguments put forward by researchers that the language system in place within the EU is itself a key component contributing to the relative success in achieving a kind of 'linguistic peace' in legal and policy discourse (Mamadouh, 1999, pp. 142) and a depoliticised equilibrium that Member States 'have little incentive to try to change' (Ringe, 2022, pp. 82), the findings of this thesis highlight the central importance of translators, terminologists and lawyer-linguists to the essence of the organisation. This thesis underlines their key role in safeguarding the ideal of equality and maintaining this equilibrium. Indeed, in this respect, the work of translators goes beyond the 'simple business of language' and instead becomes 'part of the decision-making process that eventually becomes European policy and law' (Correia, 2002, pp. 43). This research brings into sharp focus the extent to which language in the EU (including those who are responsible for creating, using and monitoring it) is inalienable from the law. Indeed, it effectively exemplifies how law is dependent on what Mattila calls 'the supernatural power of the Word' (2006, pp. 32).

This research also contributes new methodological approaches to investigating aspects of terminological variation and convergence. The methodology adopted in chapter 4 to analyse the *EUMigrLaw Corpus* took inspiration from previous research in the field of terminology (Fernández-Silva & Kerremans, 2011) in order to develop an original way of quantitatively measuring interlingual terminological convergence. This approach uses notions of terminological distance and interlingual variation index to provide comparable values on changes in the terminological 'closeness' of two

languages over time, which can then be subjected to statistical significance testing. This particular methodological approach could be replicated elsewhere and applied widely to measure the existence of terminological convergence or divergence between any two languages and provide robust data on the extent of such convergence or divergence. Moreover, chapter 6 sets out a novel methodological approach using a series of three indicators which can be combined to provide a comprehensive picture of terminological convergence or divergence across two text types. It offers a quantitative technique for measuring whether the terminology in a given field (in this case, migration) has grown more similar or more different over time in two different text types. Once again, this approach can be readily applied to any language, thematic field and text types. This research therefore provides innovative empirical approaches to evaluating terminological convergence/divergence that can then be complemented by qualitative approaches in order to offer a comprehensive insight into this phenomenon.

With regard to Migration Studies, the research set out in this thesis provides a wealth of quantitative and qualitative data and analysis demonstrating how some of the central terms in migration have developed in legal and journalistic genres in recent decades across the three languages being studied here. This research offers valuable additional empirical data for other researchers and organisations operating in the field of migration and asylum with an interest in language and terminology, and the evolving nature of terms used to refer to those involved in migration.

At this point, it is also important to address the limitations of this research project and the difficulties encountered when conducting it. Firstly, there are inevitably aspects of subjectivity inherent in the methodological approach used in the analysis of the *EUMigrLaw Corpus* with regard to determining terminological distance between the

sets of term pairs in different languages, as set out in chapter 4. Indeed, the researcher must ultimately make a judgement on each occasion regarding the extent of semantic, morphosyntactic and connotational difference between two terms when allocating a value of 0, 0.5 or 1. Since this level of subjective judgement exists in this methodological approach, it is of course possible for different researchers to obtain slightly different overall values as a result. Nevertheless, since the objective of this particular part of the methodology is to compare two time periods, the internal consistency of the study can be ensured providing the same researcher is responsible for allocating all of the values for both time periods. While not possible in this instance for reasons of resources and time, any future research using this method could further mitigate for this element of subjectivity by adopting an inter-rater agreement approach with one or more additional raters to estimate the reliability and consistency of the coding system (cf. Brezina, 2018, pp. 87).

It could also be argued that the categorisation of terms into conceptual categories involves a level of subjectivity. The researcher is required to make decisions on the breadth of the conceptual categories and where it is most appropriate to place each term that appears in the corpora. For instance, decisions needed to be made in this study such as whether '*migrant*', '*immigrant*' and '*emigrant*' should be treated as entirely separate concepts with their own categories, or whether to include them all in the broader conceptual category of MIGRANT. Indeed, as previous research involving the compiling of lexical profiles has found, 'conceptual difference may be subtle and determining when terms can be accepted to be formal variant of each other is often hard and rather like choosing a cut off point in a continuum' (Speelman, Grondelaers and Geeraerts, 2003). Moreover, much of the findings on individual terms and term pairs would remain equally valid regardless of the category

into which they were placed. In other words, the categorisation decisions do not fundamentally alter any of the key findings of this study, but simply change the way in which the study is organised and presented.

One of the main problems encountered during the study was ensuring the representativeness and balance of the *MigrationInTheNews Corpus*. A key objective at the planning stage of the study was to ensure that the English, French and Italian parts of the corpus were similar enough to allow for a meaningful terminological comparison among them. However, due to the substantially different media landscapes in the three countries, as discussed in chapter 3, it was not feasible to create a corpus with identical numbers and types of source news outlets. The print media landscape in each of the three countries has its own specific characteristics: the UK has a large proportion of tabloid newspapers, France has many widely-read weekly news magazines and Italy has a number of local or regional papers with circulations similar to those of national newspapers. In addition, the number of electronic or digitally-converted news articles available on the topic of migration varied substantially across the languages involved, especially for the earlier time period. In order to mitigate for these differences, it was nonetheless important to ensure an even coverage of differing political stances in each language and, of course, a roughly equal number of words for each language in each time period.

Also in relation to the *MigrationInTheNews Corpus*, given the significant differences in migration terminology across different types of news source within each language (e.g. between UK tabloids and broadsheets), it would be legitimate to call into question the usefulness of a group comparison given the high level of inter-textual difference within the corpus. Indeed, it is for this reason that an in-depth qualitative analysis was also carried out at this stage and the analysis was not simply limited to

quantitative analysis and statistical significance testing, which can sometimes mask specific inter-textual differences within a corpus. Although outside the scope of this particular study, it would be an interesting avenue of future study to examine the terminological differences among the individual source newspapers included in the corpus in each language. Nevertheless, it is always important to remember in any corpus-based approach that a corpus can only ever constitute a *sample* of a text type and can never capture the entire body of texts that exist within that document type or genre. The researcher's task is therefore to strive to ensure that the corpus is as representative as possible of the text type in question and that the analysis approach implemented triangulates various different methods in order to extrapolate from that an accurate picture of the linguistic phenomenon under investigation.

Having reflected on the key findings of this research study and how it advances knowledge in different academic areas, as well as its limitations, it is also interesting to consider potential directions of future research that can build on this study and practical applications it could have. From a legal perspective, although outside the scope of this particular study, its findings could be used as a basis to investigate the extent to which policy discourse shapes the language of the law. Since this thesis provides a detailed insight into the terminological dynamics in EU migration-related legislation over time, it provides an opportunity to examine possible connections between political shifts and terminological changes, exploring the degree to which terminology offers a window onto policy direction.

Indeed, language and terminology can be instrumentalised by governments and organisations responsible for developing migration policies and legislation in order to shape the way in which we think about and understand the processes and people involved (de Haas, Castles & Miller, 2020, pp. 22). The European Agenda on

Migration, produced in 2015, serves as a useful case in point here (*Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions,* 2015). The focus on ‘the plight of thousands’, ‘the human misery’ and ‘unprecedented pressure’ creates a ‘pietistic and emergency rhetoric’ (Ieracitano & Vigneri, 2018) and uses the language of humanitarianism in order to entrench draconian exceptional measures (Davitti, 2019). Meanwhile, it has also been argued that the EU’s use of euphemistic terminology in related migration legislation, such as ‘relocation’, is used to make potentially controversial expulsion policies more palatable to the public (Brannan, forthcoming, pp. 36). The language and terminology in such documents therefore offers a strong indication of the direction of migration policy within the organisation and also shows how language can be used to mask or sanitise certain measures. This also illustrates the point that many legal texts and the terminology used within them cannot be deemed to be purely legal, but are also influenced by politics and include elements belonging to areas outside the law (Mayoral Asensio, 2002). With this in mind, close monitoring of terminological changes in legal documents as showcased throughout this thesis can be harnessed to provide an insight into broader policy changes that are ongoing in a particular policy domain.

In connection with this, given the potential power of terminology in shaping perceptions and legal outcomes, the research presented in this thesis can also offer useful insights for those working to ensure consistency of terminology in EU legal documents such as the Directorate-General for Translation, the Terminology Coordination Unit (TermCoord) and the Translation Centre for the Bodies of the European Union, among others. This thesis highlights areas of variation in terminology within languages and in how terms are translated, showing the powerful

influence of source languages in shaping this terminology. As demonstrated in this thesis, this is often in direct contradiction to the very terminological resources made available to translators and drafters to guide their usage of terms in various domains. In this light, therefore, this thesis can offer valuable information on where terminology services can be improved in the interests of legal clarity and consistency in the field of migration, as well as offering methods that can be applied to other domains of competence.

The interdisciplinary research presented in this thesis has charted a course that touched on the fields of Translation Studies, Linguistics, Law and Migration Studies. By adopting a range of innovative, data-driven methods, it achieved the objectives presented at the outset and provided empirical evidence to support its conclusions. Although wide-ranging in its scope, this thesis also paves the way for future research projects exploring translation-induced language change and its effects in different fields.

References

Abdou, L. H (2020) ‘‘Push or pull’? Framing immigration in times of crisis in the European Union and the United States’, *Journal of European Integration*, 42(5), pp. 643-658

Académie Française (1999) *Questions de langue: Anglicismes et autres emprunts*. Available at: https://www.academie-francaise.fr/questions-de-langue#12_strong-em-anglicismes-et-autres-emprunts-em-strong (Accessed: 17 October 2023)

Ainsworth, J. (2014) ‘Lost in translation? Linguistic diversity and the elusive quest for plain meaning in law’ in C. Le, K. King, and A. Wagner, (eds.) *The Ashgate Handbook of Legal Translation.* Abingdon: Routledge

Aitchison, J. (2013) *Language change: Progress or decay?* (4th edition) Cambridge: Cambridge University Press

Amouzadeh, M. and House, J. (2010) ‘Translation as a language contact phenomenon. The case of English and Persian passives’, *Languages in Contrast*, 10(1), pp. 54-75

Anderman, G. and Rogers, M. (2005) ‘English in Europe: For better, for worse?’ in G. Anderman and M. Rogers (eds) *English in Europe: For better, for worse?* Clevedon: Multlingual Matters, pp. 1-26

Aprile, S. and Diaz, D. (2017) ‘Les réfugiés et l’asile dans l’Europe du XIXe siècle’ in M. Agier and Madeira, A-V. (eds) *Définir les réfugiés*, Paris: Presses Universitaires de France, pp. 29-45

Auel, K., Eisele, O. and Kinski, L. (2018) ‘What happens in parliament stays in parliament? Newspaper coverage of national parliaments in EU affairs’, *Journal of Common Market Studies*, 56(3), pp. 628-645

Aymerich, J. F., Fernández Silva, S. & Cabré Castellví, M. T. (2008) ‘La multiplicité des chemins dénominatifs’, *Meta*, 53(4), pp. 731–747

Azoulai, L. and De Vries, K. (2014) ‘Introduction’ in L. Azoulai and K. De Vries (eds) *EU Migration Law*. Oxford: Oxford University Press, pp. 1-13

Baaij, C. (2018) *Legal integration and language diversity: rethinking translation in EU lawmaking*. Oxford: Oxford University Press

Bajčić, M. (2011) ‘Conceptualization of legal terms in different fields of law: The need for a transparent terminological approach’, *Research in Language*, 9(1), pp. 81-93

Bajčić, M. (2017) *New insights into the semantics of legal concepts and the legal dictionary*. Amsterdam: John Benjamins

Baker, M. (1993) ‘Corpus Linguistics and Translation Studies – Implications and Applications’, in M. Baker, G. Francis and E. Tognini-Bonelli (eds) *Text and Technology*, Amsterdam: John Benjamins, pp. 233-252

Baker, M. (2004) ‘A corpus-based view of similarity and difference in translation’, *International Journal of Corpus Linguistics*, 9(2), pp. 167-193

Baker, M. (2020) ‘Rehumanizing the migrant: the translated past as a resource for refashioning the contemporary discourse of the (radical) left’. *Palgrave Communications,* 6(12), pp. 1-16

Baker, P., Gabrielatos, C., and McEnery, T. (2013) ‘Sketching Muslims: A corpus driven analysis of representations around the word 'Muslim' in the British Press 1998–2009’, *Applied Linguistics,* 34(3), pp. 255–278

Balič, T. (2016) ‘Attitudes towards Euro-English in a European Union Institution’, *ELOPE: English Language Overseas Perspectives and Enquiries*, 13(2), pp. 131-152

Ban, C. (2013) *Management and culture in an enlarged European Commission*. London: Palgrave Macmillan

Barnard, C. and Peers, S. (2020) *European Union Law*. 3rd edition. Oxford: Oxford University Press

Bauder, H. (2014) ‘Why we should use the term illegalized refugee or immigrant: A commentary’, International journal of refugee law, 26(3), pp. 327-332

Baumgarten, N., House, J. and Probst, J. (2004) ‘English as Lingua Franca in covert translation processes’, *The Translator*, 10(1), pp. 83-108

Becher, V., House, J. and Kranich, S. (2009) ‘Convergence and divergence of communicative norms through language contact in translation’ in K. Braunmüller and J. House (eds) *Convergence and divergence in language contact situations*. Amsterdam: John Benjamins, pp. 125-152

Bednárová-Gibová, K. (2017) ‘Translating from a lingua franca in the setting of EU translation’, *Journal of language studies, semiotics and semantics,* 8(1), pp.148-157

Bell, A. (1991) *Language of the news media*. Oxford: Blackwell

Bennett, K. (2011) ‘The scientific revolution and its repercussions on the translation of technical discourse’, *The Translator*, 17(2), pp. 189-210

Berenskoetter, F. (2016) ‘Unpacking concepts’ in F. Berenskoetter (ed) *Concepts in world politics*, London: Sage Publications, pp. 1-22

Berkeley, R., Khan, O. and Ambikaipaker, M. (2006) *What’s new about new immigrants in twenty-first century Britain?* York: Joseph Rowntree Foundation

Bernardini, S. and Ferraresi, A. (2011) ‘Practice, description and theory come together – Normalization or interference in Italian technical translation?’, *Meta: Journal des traducteurs*, 56(2), pp. 226-246

Berry, M., Garcia-Blanco, I., and Moore, K. (2015) *Press coverage of the refugee and migrant crisis in the EU: A content analysis of five European countries*. [Project Report]. Geneva: United Nations High Commissioner for Refugees. Available at: http://www.unhcr.org/56bb369c9.html (Accessed: 12 September 2023)

Bertaccini, F., Massari, M. and Castagnoli, S. (2010) ‘Synonymy and variation in the domain of digital terrestrial television: Is Italian at risk?’ in M. Thelen and F. Steurs (eds) *Terminology in everyday life*. Amsterdam: John Benjamins

Biber, D. (2009) ‘Corpus-based and corpus-driven analyses of language variation and use’ in *The Oxford handbook of linguistic analysis (1st edition),* Oxford: Oxford University Press

Biber, D. (2015) ‘Corpus-based and corpus-driven analyses of language variation and use’ in *The Oxford handbook of linguistic analysis (2nd edition),* Oxford: Oxford University Press

Biel, Ł. (2014) *Lost in the Eurofog: The textual fit of translated law.* Frankfurt am Main: Peter Lang

Biel, Ł. (2019) ‘Theoretical and methodological challenges in researching EU legal translation’ in I. Simonnaes and M. Kritiansen (eds) *Legal translation. current issues and challenges in research, methods and application.* Berlin: Frank & Timme

Biel, Ł., Biernacka, A., and Jopek-Bosiacka, A. (2018) ‘Collocations of terms in EU competition law: A corpus analysis of EU English collocations: The role of language and translation in EU competition law’ in S. Marino, Ł. Biel, M. Bajčić and V. Sosoni (eds) *Language and law. The role of language and translation in EU competition law*. Cham: Springer, pp. 249-274

Biel, Ł. and Koźbiał, D. (2020) ‘How do translators handle (near-) synonymous legal terms? A mixed-genre parallel corpus study into the variation of EU English-Polish competition law terminology’, *Estudios de Traducción*, 10, pp. 69-90

Bisiada, M. (2013) ‘Changing conventions in German causal clause complexes: A diachronic corpus study of translated and non-translated business articles’, *Languages in Contrast*, 13(1), pp. 1-27

Bisiada, M. (2016) ‘Structural effects of English–German language contact in translation on concessive constructions in business articles’, *Text & Talk*, 36(2), pp.133-154

Blinder, S. and Allen, W. L. (2016) ‘Constructing immigrants: portrayals of migrant groups in British national newspapers, 2010-2012’, *International migration review*, 50(1), pp.3-40

Boswell, C. (2014) ‘The politics of irregular migration’ in L. Azoulai and K. De Vries (eds) *EU migration law*. Oxford: Oxford University Press, pp. 41-68

Brannan, J. (forthcoming) ‘From the domestic to the supranational: the terminology of “expulsion” as used at the European Court of Human Rights’ in Ł. Biel (ed) *Benjamins handbook of legal terminology*. Amsterdam: John Benjamins

Brezina, V. (2018) *Statistics in corpus linguistics. A practical guide*. Cambridge: Cambridge University Press

Brouwer, E. (2003) ‘Immigration, asylum and terrorism: A changing dynamic legal and practical developments in the EU in response to the terrorist attacks of 11.09’, *European journal of migration and law*, 4, pp. 399-424

Brownlie, S. (2010) ‘Representing news from France’ in C. Schäffner and S. Bassnett (eds) *Political discourse, media and translation.* Cambridge: Cambridge Scholars Publishing, pp. 32-54

Cao, D. (2007) *Translating law*. Clevedon: Multilingual Matters

Carrera, S. (2014) ‘Integration of immigrants in EU law and policy’ in L. Azoulai and K. De Vries (eds) *EU migration law*. Oxford: Oxford University Press, pp. 149-187

Čavoški, A. (2017) ‘Interaction of law and language in the EU: Challenges of translating in multilingual environment’, *Journal of specialised translation*, 27, pp. 58-74

Čavoški, A. (2018) ‘Legal language and EU integration. The case of the Western Balkans’, *International journal of language & law*, 7, pp. 70-96

Čavoški, A. (2020) ‘Science and law in environmental law and policy: The case of the European Commission’ in *Transnational environmental law*, 9(2), pp. 263-295

Charteris-Black, J. (2006) ‘Britain as a container: Immigration metaphors in the 2005 election campaign’, *Discourse & society*, 17(5), pp. 563-581

Clarke, H. D., Goodwin, M. and Whiteley, P. (2017) *Brexit: why Britain voted to leave the European Union*. Cambridge: Cambridge University Press

Clay, E. (2022) 'A corpus-based approach to examining terminological variation in EU law', *International journal of language and law*, 11, pp. 142-162

Clay, E. (forthcoming) 'Language contact within an institutional ecosystem: The impact of EU translation' in Cerarols, R. and Dasca, M. (eds) *Translation studies and ecology. Mapping the possibilities of a new emerging field*. Abingdon: Routledge

Clay, E. and McAuliffe, K. (2021) 'Reconceptualising the third space of legal translation: A study of the Court of Justice of the European Union', *Comparative legilinguistics*, 45(1), pp. 93-126

Collett, E. and Le Coz, C. (2018) *After the storm: Learning from the EU response to the migration crisis*. Brussels: Migration Policy Institute Europe

Cornu, G. (2020) *Vocabulaire Juridique*. Paris: Presses Universitaires de France.

Correia, R. (2002) 'Translation of EU legal texts' in A. Tosi (ed) *Crossing barriers and bridging cultures*. Clevedon: Multilingual Matters

Crawley, H. and Skleparis, D. (2018) 'Refugees, migrants, neither, both: categorical fetishism and the politics of bounding in Europe's 'migration crisis'', *Journal of ethnic and migration studies*, 44(1), pp. 48-64

Crépeau, F. (2013) *UN Special Rapporteur on the human rights of migrants concluding remarks*, [Speech], FRA-ECtHR seminar on European law on asylum, borders and immigration, Strasbourg, 11 June

Dal Monte, P. (2018) *Migranti, rifugiati, sfollati, profughi: mettiamo ordine nel linguaggio politicamente corretto.* Available at: https://www.ilfattoquotidiano.it/2018/11/30/migranti-rifugiati-sfollati-profughi-

mettiamo-ordine-nel-linguaggio-politicamente-corretto/4800886/ (Accessed: 11 September 2023)

Dauvergne, C. (2008) *Making people illegal: What globalization means for migration and law*. Cambridge: Cambridge University Press

Davitti, D. (2019) 'The rise of private military and security companies in European Union migration policies: Implications under the UNGPs', *Business and human rights journal*, 4(1), pp. 33-55

De Bessé, B., Nkwenti-Azeh, B. and Sager. J. C. (1997) 'Glossary of terms used in terminology', *Terminology*, 4(1), pp. 117-156

De Haas, H, Castles, S. and Miller, M. J. (2020) *The age of migration: international population movements in the modern world*. 6th edition. London: Red Globe Press

De Swaan, A. (2001) *Words of the world*. Cambridge: Polity

De Vreese, C. H. (2005) 'News framing: Theory and typology', *Information design journal*, 13(1), pp. 51-62

Dijkstra, L., Poelman, H. and Rodríguez-Pose, A. (2020) 'The geography of EU discontent', *Regional Studies*, 54(6), pp. 737-753

Directorate-General for migration and home affairs (2020) *European migration network*. Available at: https://home-affairs.ec.europa.eu/networks/european-migration-network-emn_en. (Accessed: 3 October 2023)

Doczekalska, A. (2009) 'Drafting and interpretation of EU law – paradoxes of EU multilingualism' in Grewendorf, G. and Rathert, M. (eds.) *Formal linguistics and law.* Berlin: De Gruyter Mouton

Dollerup, C. (2004) ‘The vanishing original’, *Hermes, journal of linguistics*, 32, pp. 185-199

Edwards, A. (2016) *UNHCR viewpoint: ‘Refugee’ or ‘Migrant’ – which is right?* Available at: https://www.unhcr.org/news/stories/unhcr-viewpoint-refugee-or-migrant-which-right#:~:text=We%20say%20'refugees'%20when%20we,Choices%20about%20words%20do%20matter (Accessed: 14 September 2023)

EU monitor (2010) *Legal instruments.* Available at: https://www.eumonitor.eu/9353000/1/j9vvik7m1c3gyxp/vh75mdhkg4s0 (Accessed: 28 September 2023)

EU monitor (2010) *Protocol*. Available at: https://www.eumonitor.eu/9353000/1/j9vvik7m1c3gyxp/vh7gkw69l2xw. (Accessed: 3 October 2023)

European Commission (2015) *DGT translation quality guidelines*. Available at: https://ec.europa.eu/translation/maltese/guidelines/documents/dgt_translation_quality_guidelines_en.pdf (Accessed: 14 September 2023)

European Commission (2017) *Asylum quarterly report*. Available at: https://ec.europa.eu/eurostat/documents/12544011/12980988/Asylum+Quarterly+Report+-+Q4+2017.pdf/155662e9-72df-b592-4e46-3d0dd5fee6ef?t=1625511347269 (Accessed: 19 October 2023)

European e-Justice (2019) *EU law*. Available at: https://e-justice.europa.eu/content_eu_law-3-en.do (Accessed: 3 October 2023)

European Union (2015) *Joint practical guide of the European parliament, the Council and the Commission for persons involved in the drafting of European Union legislation*. Luxembourg: Publications office of the European Union. Available at: https://eur-lex.europa.eu/content/techleg/KB0213228ENN.pdf. (Accessed: 30 October 2023)

European Union (2019) *Standard Eurobarometer 92*. Available at: https://europa.eu/eurobarometer/surveys/detail/2255 (Accessed: 24 October 2023)

Faber, P. and León-Araúz, P. (2016) 'Specialized knowledge representation and the parameterization of context', *Frontiers in psychology*, 7(196), pp. 1-20

Farkas, A. (2010) *LF Aligner*. [Computer programme] Available at: https://sourceforge.net/projects/aligner (Accessed: 3 October 2023)

Felber, H. (1984) *Terminology manual*. Paris: UNESCO and Infoterm

Felici, A. (2010) 'Translating EU law: legal issues and multiple dynamics', *Perspectives: Studies in translatology*, 18(2), pp. 95-108

Fernández-Silva, S. & Kerremans, K. (2011) 'Terminological variation in source texts and translations: A pilot study', *Meta*, 56(2), pp. 318–335

Filipovic, L. (2013) 'The role of language in legal contexts: A forensic cross-linguistic viewpoint' in M. Freeman and F. Smith (eds.) *Law and language: Current legal issues, Volume 15*. Oxford: Oxford University Press

Firmstone, J. (2018) *United Kingdom – media landscapes*. Available at: https://medialandscapes.org/country/united-kingdom (Accessed: 3 October 2023)

Fischer, M. (2009) ‘National and European language policies – case study Hungary’ in G. Stickel (ed) *National and European language policies. Contributions to the annual conference 2008 of EFNIL in Riga*. Bern: Peter Lang

Fischer, M. (2010) ‘Language (policy), translation and terminology in the European Union’ in M. Thelen and F. Steurs (eds) *Terminology in everyday life*. Amsterdam: John Benjamins’

Follesdal, A. and Hix, S. (2006) ‘Why there is a democratic deficit in the EU: A response to Majone and Moravcsik’, *Journal of common market studies*, 44(3), pp. 533-562

Fotopoulos, S. and Kaimaklioti, M. (2016) ‘Media discourse on the refugee crisis: On what have the Greek, German and British press focused?’, *European view*, 15, pp. 265-279

Fowler, R. (1991) *Language in the news: Discourse and ideology in the press*. Abingdon: Routledge

Freixa, J. (2006) ‘Denominative variation in terminology’, *Terminology*, 12(1), pp. 51-77

Gatrell, P. (2019) *The unsettling of Europe: How migration reshaped a continent*. New York: Basic Books

Geeraerts, D. (2010) *Theories in lexical semantics*. Oxford: Oxford University Press

Goddard, C. and Wierzbicka, A. (2014) *Words & meanings: Lexical semantics across domains, languages & cultures*. Oxford: Oxford University Press

Goffin, R. (1994) ‘L’eurolecte: Oui, jargon communautaire: non’, *Meta: Journal des traducteurs*, 39(4), pp. 636-642

Greco. A. (2022) *Migranti, profughi, richiedenti asilo e rifugiati: quali differenze*. Available at: https://www.laleggepertutti.it/557020_migranti-profughi-richiedenti-asilo-e-rifugiati-quali-differenze (Accessed: 3 October 2023)

Greussing, E. and Boomgaarden, H. J. (2017) 'Shifting the refugee narrative? An automated frame analysis of Europe's 2015 refugee crisis', *Journal of ethnic and migration studies*, 43(11), pp. 1749-1774

Guia, M. J. (2016) *Building ideologies: bias in the legal translation of irregularity/Illegality terminology*. Master's thesis. University of Coimbra

Hansen-Schirra, S. (2011) 'Between normalization and shining-through: Specific properties of English-German translations and their influence on the target language' in S. Kranich, V. Becher, S. Höder and J. House (eds) *Multilingual discourse production. Diachronic and synchronic perspectives*. Amsterdam: John Benjamins pp. 133-162

Heine, B. and Kuteva, T. (2005) *Language contact and grammatical change*. Cambridge: Cambridge University Press

Home Office (2022) *Foreign criminals and illegal entrants to the UK removed to Albania*. Available at: https://www.gov.uk/government/news/foreign-criminals-and-illegal-entrants-to-the-uk-removed-to-albania--2. (Accessed: 24 October 2023)

Hoops, J. F. and Braitman, K. (2019) 'The influence of immigration terminology on attribution and empathy', *Critical discourse studies*, 16(2), pp. 149-161

House, J. (1997) *Translation quality assessment. A model revisited*. Tübingen: Narr

House, J. (2006) 'Covert translation, language contact, variation and change', *SYNAPS*, 19, pp. 25-47

House, J. (2011) ‘Linking constructions in English and German translated and original texts’ in S. Kranich, V. Becher, S. Höder and J. House (eds) *Multilingual discourse production. Diachronic and synchronic perspectives*. Amsterdam: John Benjamins, pp. 163-182

IBM (2022) *SPSS Statistics* (Version 29) [Computer programme]. Available at: https://www.ibm.com/products/spss-statistics (Accessed: 10 October 2023)

Ieracitano, F. and Vigneri, F. (2018) ‘In "their" words and in "our" words: a comparison between European policies, media narratives and migrants’ testimonies of landings in the Mediterranean’, *Language, discourse and society*, 6(1), pp. 62-82

ILO (International Labour Organization) (2020) *Media-friendly glossary on migration*. Available at: https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/documents/publication/wcms_767998.pdf. (Accessed: 17 October 2023)

ILSP & MCRC (2000) *The Treaties*. Available at: http://www.xanthi.ilsp.gr/kemeseu/ch1/treaties.htm (Accessed: 3 October 2023)

IOM (International Organization for Migration) (2015) *Irregular Migrant, Refugee Arrivals in Europe Top One Million in 2015: IOM*. Available at: https://www.iom.int/news/irregular-migrant-refugee-arrivals-europe-top-one-million-2015-iom (Accessed: 19 October 2023)

IOM (International Organization for Migration) (2022) *Key migration terms*. Available at: https://www.iom.int/key-migration-terms# (Accessed: 17 October 2023)

IRC (International Rescue Committee) (2022) *IRC Italy*. Available at: https://www.rescue.org/country/italy (Accessed: 19 October 2023)

ISMU – Iniziative e studi sulla multietnicità (2006) *L'extracomunitario: definizione e status*. Available at: https://archive.ph/20120906054649/http:/www.ismu.org/approfondimento.php?id=167 (Accessed: 24 October 2023)

Johanson, L. (1993) 'Code-copying in immigrant Turkish', in G. Extra and L. Verhoeven (eds) *Immigrant languages in Europe*, Clevedon: Multilingual Matters, pp. 197-221

Johanson, L. (1999) 'The dynamics of code-copying in language encounters', in B. Brendemoen, E. Lanza, and E. Ryen (eds) *Language encounters across time and space,* Oslo: Novus Press, pp. 37-61

Johanson, L. (2002) 'Contact-induced change in a code-copying framework' in M. C. Jones and E. Esch (eds) *Language change: The interplay of internal, external and extra-linguistic factors*, Berlin: Mouton De Gruyter, pp. 285-313

Kaguera, K. (2002) *The dynamics of terminology: A descriptive theory of term formation and terminological growth*. Amsterdam: John Benjamins

Kardos, O. H. (2023) 'The role of EU translators in countering domain loss. A case study'. *Terminology: Domain loss and gain*. Conference held at KU Leuven, Brussels Campus. 20-21 April

Karyotis, G. (2007) 'European migration policy in the aftermath of September 11', *Innovation: The European journal of social science research*, 20(1), pp. 1-17

Kenning, M-M. (2010) 'What are parallel and comparable corpora and how can we use them?' in A. O'Keeffe and M. McCarthy (eds) *The Routledge handbook of corpus linguistics*. Abingdon: Routledge, pp. 487-500

Kermas, S. (2010) ‘English legal discourse and the French continuum’ in D. S. Giannoni and C. Frade (eds) *Researching language and the law: Textual features and translation issues.* Frankfurt am Main: Peter Lang, pp. 19-48

Kerremans, K. and Temmerman, R. (2016) ‘How terminological equivalence differs from translation equivalence: Quantitative and qualitative comparisons of term variants and their translations in a parallel corpus of EU texts’, in G. Corpas Pastor and M. Seghiri (eds) *Corpus-based approaches to translation and interpreting: From theory to applications*, Frankfurt am Main: Peter Lang, pp. 43-64

Khosravinik, M. (2009) ‘The representation of refugees, asylum seekers and immigrants in British newspapers during the Balkan conflict (1999) and the British general election (2005)’, *Discourse and Society*, 20(4), pp. 477-498

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014) ‘The Sketch Engine: Ten years on’, *Lexicography*, 1, pp. 7-36

Kolehmainen, L. and Riionheimo, H. (2016) Literary translation as language contact: A pilot study on the Finnish passive, *International journal of literary linguistics*, 5(3), pp. 1-32

Koskinen, K. (2000) ‘Institutional Illusions’, *The Translator*, 6(1), pp. 49-65

Koskinen, K. (2008) *Translating institutions: An ethnographic study of EU translation*. New York: St. Jerome Publishing

Kostas, S. (2017) ‘Irregular vs. Illegal immigration: Setting the definitions. An overview of European practice’, *Slovenský národopis*, 65(4), pp. 420–426

Koutonin, M. R. (2015) *Why are white people expats when the rest of us are immigrants?* Available at: https://www.theguardian.com/global-development-professionals-network/2015/mar/13/white-people-expats-immigrants-migration (Accessed: 11 September 2023)

Kranich, S. (2014) 'Translations as a locus of language contact' in J. House (ed) *Translation: A multidisciplinary approach*. London: Palgrave Macmillan

Kranich, S., Becher, V. and Höder, S. (2011) 'A tentative typology of translation-induced language change' in S. Kranich, V. Becher, S. Höder and J. House (eds) *Multilingual discourse production: Diachronic and synchronic perspectives*, Amsterdam: John Benjamins, pp. 9-44

Krzyżanowski, M., Triandafyllidou, A. & Wodak, R. (2018) 'The mediatization and the politicization of the "refugee crisis" in Europe', *Journal of immigrant & refugee studies*, 16(1-2), pp. 1-14

Labov, W. (1981) 'What can be learned about change in progress from synchronic description?' in D. Sankoff and H. Cedergreen (eds) *Variation omnibus*. Edmonton: Linguistic Research, pp. 177-200

Lakoff, G. and Ferguson, S. (2006) 'The framing of immigration', *UC Berkeley*. Available at: https://escholarship.org/uc/item/0j89f85g. (Accessed: 14 September 2023)

Lardeau, M. (2018) *France – media landscapes*. Available at: https://medialandscapes.org/country/france. (Accessed: 3 October 2023)

Laviosa, S. (2006) ‘Data-driven learning for translating anglicisms in business communication’, *IEEE transactions on professional communication*, 49(3), pp. 267-274

Leech, G. (1991) ‘The state of the art in corpus linguistics’ in K. Aijmer and B. Altenberg (eds) *English corpus linguistics: studies in honour of Jan Svartvik*. London: Longman

Legrand, P. (2005) ‘Issues on the translatability of law’ in Bermann, S. and Wood, M. (eds.) *Nation, language and the ethics of translation*. Princeton: Princeton University Press.

León-Araúz, P., Cabezas-García, M. and Reimerink, A. (2020) ‘Representing multiword term Variation in a terminological knowledge base: A corpus-based study’, *Proceedings of the 12th conference on language resources and evaluation (LREC 2020)*, Marseille, 11-16 May 2020. European language resources association, pp. 2358–2367

Lequesne, C. (2020) *Diversité linguistique et langue française en Europe*. Ateliers Henry Dougier. Available at: https://www.vie-publique.fr/rapport/282161-diversite-linguistique-et-langue-francaise-en-europe (Accessed: 13 September 2023)

Leung, J. (2014) ‘Translation equivalence as legal fiction’ in C. Le, K. King, and A. Wagner (eds.) *The Ashgate handbook of legal translation.* Abingdon: Routledge.

Liao, M-H. (2010) ‘Influence of translations on non-translations: Popular science as a new genre in Taiwan’, *Perspectives: Studies in translatology*, 18(2), pp. 127-140

Loupaki, E. (2018) ‘EU legal language and translation – Dehumanizing the refugee crisis’, *International journal of language and law*, 7, pp. 97-116

Maas, W. (2007) *Creating European citizens*. Milton Keynes: Rowman and Littlefield

Mahrouse, G. (2018) ‘The construction of European vulnerability in the Guardian news headlines on Calais: Portraying migrants as threats and moderating the dangers they face’, *Language, discourse and society*, 6(1), pp. 32-42

Malamatidou, S. (2011) ‘Translation and language change with reference to popular science articles: The interplay of diachronic and synchronic corpus-based studies’ in H. Hedeland, T. Schmidt and K. Worner (eds), *Multilingual resources and multilingual applications: Proceedings of the conference of the German society for computational linguistics and language technology (GSCL) 2011*. *vol. 96*, Hamburg: University of Hamburg, pp. 135-140

Malamatidou, S. (2016) 'Understanding translation as a site of language contact: The potential of the code-copying framework as a descriptive mechanism in translation studies', *Target*, 28(3), pp. 399-423

Malone, B. (2015) *Why Al Jazeera will not say Mediterranean ‘migrants’*. Available at: https://www.aljazeera.com/features/2015/8/20/why-al-jazeera-will-not-say-mediterranean-migrants. (Accessed: 14 September 2023)

Mamadouh, V. (1999) ‘Beyond nationalism: Three visions of the European Union and their implications for the linguistic regime of its institutions’, *GeoJournal*, 48, pp. 133-144

Mancini, P and Gerli. M. (2018) *Italy – media landscapes*. Available at: https://medialandscapes.org/country/italy/media/print (Accessed: 3 October 2023)

Mariani, J. (2018) *Migration in translation: The role of terminology and trans-editing in shaping the crisis in EU institutions*. PhD Thesis. University of Verona. Available

at: https://www.dlls.univr.it/?ent=pubbdip&id=981576. (Accessed: 14 September 2023)

Mariani, J. (2021) 'Migration terminology in the EU Institutions', *Terminology*, 27(1), pp. 35-55

Marsh, D. (2015) *We deride them as 'migrants'. Why not call them people?* Available at: https://www.theguardian.com/commentisfree/2015/aug/28/migrants-people-refugees-humanity (Accessed: 14 September 2023)

Martínez Lirola, M. (2021) 'Approaching the discourse of hospitality and solidarity regarding the rescue of the ship Aquarius in a sample of the Spanish press', *Fonseca journal of communication*, 22, pp. 189-205

Matilla, H. (2006) *Comparative legal linguistics*. Farnham: Ashgate Publishing

Mayoral Asensio, R. (2002) 'Cómo se hace la traducción jurídica?', *Puentes*, 2, pp. 9-14

McAuliffe, K. (2011) 'Hybrid texts and uniform law? The multilingual case law of the Court of Justice of the European Union', *International Journal of Semiotics of Law*, 24, pp. 97-115

McAuliffe, K. (2016) 'Hidden translators: the invisibility of translators and the influence of lawyer-linguists on the case law of the Court of Justice of the European Union', *Language and Law / Linguagem e Direito*, 3(1), pp. 5-29

McEnery, T. and Wilson, A. (2001) *Corpus linguistics*. 2nd edition. Edinburgh: Edinburgh University Press

McLaughlin, M. (2010) 'L'influence de l'anglais sur la syntaxe du français: une étude de cas concernant la voix passive' in F. Neveu, V. Muni Toke, J. Durand, T. Klingler,

L. Mondada and S. Prévost (eds.) *Congrès Mondial de Linguistique Française - CMLF 2010*, Paris: Institut de Linguistique Française

McLaughlin, M. (2013) 'News translation as a source of syntactic borrowing in Italian', *The Italianist*, 33(3), pp. 443-463

Meyer, I. and Mackintosh, K. (2000) 'When terms move into our everyday lives: An overview of de-terminologization', *Terminology*, 6(1), pp. 111-138

Mori, L. (2018a) 'Introduction: The Eurolect observatory project' in L. Mori (ed.) *Observing Eurolects: Corpus analysis of linguistic variation in EU law*. Amsterdam: John Benjamins

Mori, L. (2018b) 'Observing Eurolects: The case of Italian' in L. Mori (ed.) *Observing Eurolects: Corpus analysis of linguistic variation in EU law*. Amsterdam: John Benjamins

Mori, L. (2018c) 'Conclusions: A cross-linguistic overview on Eurolects' in L. Mori (ed) *Observing Eurolects: Corpus analysis of linguistic variation in EU law*. Amsterdam: John Benjamins

Mould. M. (2011) *The Routledge dictionary of cultural references in modern French*. London: Routledge

Musacchio, M. T. (2005) 'The influence of English on Italian: The case of translations of economics articles' in G. Anderman and M. Rogers (eds) *In and out of English: For better, for worse*. Bristol: Multilingual Matters, pp.71-96

Nelken, G. (2018) 'The social construction of asylum seekers in the refugee crisis: hyper-coverage without agency', *Language, discourse and society*, 6(1), pp. 43-61

Nelson, M. (2010) ‘Building a written corpus: what are the basics?’ in A. O’Keeffe and M. McCarthy (eds) *The Routledge handbook of corpus linguistics*. Abingdon: Routledge, pp. 53-65

Neumann, S. (2011) ‘Assessing the impact of translations on English-German language contact: Some methodological considerations’, in in S. Kranich, V. Becher, S. Höder and J. House (eds) *Multilingual discourse production: Diachronic and synchronic perspectives*, Amsterdam: John Benjamins, pp. 233-255

Olsen, N. (2012) *History in the plural: An introduction to the work of Reinhart Koselleck*. Oxford: Berghahn Books

Otero Fernández, I. (2020) *Multilingualism and the meaning of EU law*. PhD Thesis. European University Institute. Available at: https://cadmus.eui.eu/handle/1814/66308 (Accessed: 13 September 2023)

Pace, P. and Severance, K. (2016) ‘Migration terminology matters’, *Forced migration review*, 51, pp. 69-70

Pang, S. and Wang, K. (2020) ‘Language contact through translation: The influence of explicitness in English-Chinese translation on language change in vernacular Chinese’, *Target*, 32(3), pp. 420-455

Pearson, J. (1998) *Terms in context*. Amsterdam: John Benjamins

Peers, S. (2011) *EU justice and home affairs law*. 3rd edition. Oxford: Oxford University Press

Pérez-Paredes, P., Jiménez P. & Hernández, P. (2017) ‘Constructing immigrants in UK legislation and administration informative texts: A corpus-driven study (2007–2011)’, *Discourse and society*, 28(1), pp. 81-103

Perkowska, M. (2016) ‘Illegal, legal, irregular or regular – Who is the incoming foreigner?’, *Studies in logic, grammar and rhetoric*, 45(58), pp. 187-197

Peruzzo, K. (2013) ‘Short-period evolution in EU legal texts: old and new terms, old and new concepts’, *Linguistica*, 53(2), pp. 39-53

Picton, A. (2014) ‘The dynamics of terminology in short-term diachrony’ in R. Temmerman and M. Van Campenhoudt (eds) *Dynamics and terminology. An interdisciplinary perspective on monolingual and multilingual culture-bound communication*, Amsterdam: John Benjamins, pp. 159-182

Pavliuk, N. (2023) ‘A potential of legal terminology to be translated: The case of ‘*regulation*’ translated into Ukrainian’. *International journal for the semiotics of law*

Prieto Ramos, F. (2019) ‘The use of corpora in legal and institutional translation studies’, *Translation spaces*, 8(1), pp. 1-11

Prieto Ramos, F. and Guzman, D. (2023) ‘Translating legal terms at international organisations: Do institutional term banks meet translators’ needs?’ in J. Zhao and V. L. C. Lei (eds), *New advances in legal translation and interpreting, new frontiers in translation studies*. Singapore: Springer, pp. 243-264

Prieto Ramos, F. and Morales Moreno, A. (2019) ‘Terminological innovation and harmonization at international organizations: Can too many cooks spoil the broth?’ in Simonnaes, I. and Kristiansen, M. (eds) *Legal translation. Current issues and challenges in research, methods and applications,* Berlin: Frank & Timme, pp. 87-112

Publications Office of the European Union (2022) *Interinstitutional style guide*, Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2830/215072 (Accessed: 14 September 2023)

Rayson, P. (2016) *Log-likelihood and effect size calculator*. Available at: https://ucrel.lancs.ac.uk/llwizard.html (Accessed: 19 October 2023)

Rayson, P. and Garside, R. (2000). 'Comparing corpora using frequency profiling' in *WCC '00 Proceedings of the workshop on comparing corpora.* vol. 9, pp. 1-6, Workshop on comparing corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000), Hong Kong

RELX (2023) *Lexis Nexis*. Available at: https://advance.lexis.com/BisNexisNewsSearch/?pdmfid=1519360&crid=adb53914-87f2-475c-9ee4-1078c718982b (Accessed: 3 October 2023)

Resi, R. and ten Hacken, P. (2023) 'Perdita di domini: Is Italian losing domains?' *Terminology: Domain loss and gain*. Conference held at KU Leuven, Brussels Campus. 20-21 April

Ringe, N. (2022) *The language(s) of politics. Multilingual policy-making in the European Union*. Ann Arbor: University of Michigan Press

Robertson, C. (2010) 'EU Law and Semiotics', *International journal for the semiotics of law*, 23(2), pp. 145-165

Robertson, C. (2011) 'Multilingual legislation in the European Union: EU and national legislative-language styles and terminology', *Research in language*, 9(1), pp. 51-67

Robertson, C. (2014) 'EU legislative texts and translation' in Le, C., King, K., and Wagner, A. (eds.) *The Ashgate handbook of legal translation.* Abingdon: Routledge

Robinson, M. (2018) *A man needs a female like a fish needs a lobotomy: The role of adjectival nominalization in pejorative meaning*. Master's Thesis. University of North Texas. Available at; https://digital.library.unt.edu/ark:/67531/metadc1157617/m2/1/high_res_d/ROBINSON-THESIS-2018.pdf (Accessed: 24 October 2023)

Robinson, W. (2014) 'Translating legislation: The European Union experience', *The theory and practice of legislation*, 2(2), pp. 185-210

Rogato, G. (2014) 'Anglicismi nella stampa italiana', *Italica*, 85(1), pp. 27-43

Rollason, C. (2005) 'Unequal systems: on the problems of anglicisms in contemporary French usage' in G. Anderman and M. Rogers (eds) *English in Europe: For better, for worse?* Clevedon: Multlingual Matters, pp. 39-56

Russo Bullaro, G. (2010) *From Terrone to Extracomunitario: New manifestations of racism in contemporary Italian cinema*. Market Harborough: Troubador Publishing

Sager, J. C. (1997) 'Term formation' in S. E. Wright and G. Budin (eds) *Handbook of terminology management. Volume 1: Basic aspects of terminology management*. Amsterdam: John Benjamins, pp. 25-41

Šarčević, S. (1997) *New approach to legal translation*. The Hague: Kluwer Law International.

Šarčević, S. (2013) 'Multilingual lawmaking and legal (un)certainty in the European Union', *International journal of law, language and discourse*, 3(1), pp. 1-29

Scotto, A. (2017) *From emigration to asylum destination, Italy navigates shifting migration tides*. Available at: https://www.migrationpolicy.org/article/emigration-

asylum-destination-italy-navigates-shifting-migration-tides (Accessed: 19 October 2023)

Serlet, K. (2023) ‘EU institutions’ linguistic services – a catalyst for terminology development at national level’, *Terminology: Domain loss and gain*. Conference held at KU Leuven, Brussels Campus. 20-21 April

Shehata, A. and Strömbäck, J. (2011) 'A matter of context: A comparative study of media environments and news consumption gaps in Europe', *Political communication*, 28(1), pp. 110-134

Sigona, N. (2018) ‘The contested politics of naming in Europe’s “refugee crisis”’, *Ethnic and racial studies*, 41(3), pp. 456-460

Sirhan. N. R. (2021) *Reporting Palestine-Israel in British newspapers*. Cham: Palgrave Macmillan

Solan, L. (2014) Multilingualism and morality in statutory interpretation, *Language and Law/Linguagem e Direito*, 1(1), pp. 5-21.

Somerville, W. (2007) *Immigration under New Labour*. Bristol: Bristol University Press

Sosoni, V. (2012) A hybrid translation theory for EU texts, *Vertimo Studijos*, 50, pp. 76-89

Speelman, D., Grondelaers, S. and Geeraerts, D. (2003) ‘Profile-based linguistic uniformity as a generic method for comparing language varieties’, *Computers and the humanities*, 37, pp. 317-337

Steiner, E. (2008) ‘Empirical studies of translations as a mode of language contact – “explicitness” of lexicogrammatical encoding as a relevant dimension’ in P. Siemund

and N. Kintana (eds) *Language contact and contact languages*. Amsterdam: John Benjamins, pp. 317-341

Stewart, D. (2010) *Semantic prosody: A critical evaluation*. Abingdon: Routledge

Strandvik, I. (2012) 'Legal harmonization through legal translation: Texts that say the same thing?' in C. J. W. Baaij (ed) *The role of legal translation in legal harmonization*. Alphen aan den Rijn: Wolters Kluwer, pp. 25-49

Stumpf, J. (2006). 'The crimmigration crisis: Immigrants, crime, and sovereign power', *American University Law Review*, 56(2), pp. 367-420

Taylor, C. (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), pp. 368-400

Taylor, C. (2021) 'Metaphors of migration over time', *Discourse and society*, 32(4), pp. 1-19

Teich, E. (2003) *Cross-linguistic variation in system and text. A methodology for the investigation of translations and comparative texts*. The Hague: De Gruyter Mouton

Temmerman, R. (2000) *Towards new ways of terminology description. The sociocognitive approach*. Amsterdam: John Benjamins

Temmerman, R. (2018) 'European Union multilingual primary term creation and the impact of its neologisms on national adaptations', *Parallèles*, 30(1), pp. 8-20

Terminology Coordination Unit (2018) *What we do*. Available at: https://termcoord.eu/what-we-do/ (Accessed: 17 October 2023)

Thomason, S. (2001) *Language contact: An introduction*. Edinburgh: Edinburgh University Press.

Thomason, S. (2006) 'Language change and language contact' in K. Brown (ed) *Encyclopedia of language and linguistics* (2nd edition). Amsterdam: Elsevier

Thomason, S. (2020) 'Contact explanations in linguistics', in R. Hickey (ed) *The handbook of language contact* (2nd edition). Chichester: John Wiley & Sons

Thomason, S. and Kaufman, T. (1988) *Language contact, creolization and genetic linguistics*. Oakland: University of California Press

Tiersma, P. (1999) *Legal language*. Chicago: University of Chicago Press

Toner, H. (2014) 'The Lisbon Treaty and the future of European immigration and asylum law', in L. Azoulai and K. De Vries (eds) *EU Migration Law*. Oxford: Oxford University Press, pp. 14-40

Tosi, A. (2007) *Un Italiano per l'Europa: La traduzione come prova di vitalità*. Rome: Carocci editore.

Toury, G. (1995) *Descriptive translation studies – and beyond*. Amsterdam: John Benjamins

Treccani (2022) *Extracomunitario*. Available at: https://www.treccani.it/vocabolario/extracomunitario/ (Accessed: 24 October 2023)

Triandafyllidou, A. (2014) 'Greek migration policy in the 2010s: Europeanization tensions at a time of crisis', *Journal of European integration*, 36(4), pp. 409-425

Trim, J. (2002) ‘Multilingualism and the interpretation of languages in contact’ in Tosi, A. (ed) *Crossing barriers and bridging cultures*. Clevedon: Multilingual Matters, pp. 8-20

Trklja, A. and McAuliffe, K. (2020) ‘Formulaic metadiscursive signalling devices in judgments of the Court of Justice of the European Union: A new corpus-based model for studying discourse relations of texts’, *International journal of speech, language and the law*, 26(1), pp. 21-55

UNODC (United Nations Office on Drugs and Crime) (2018) *Human trafficking and migrant smuggling.* Available at: https://www.unodc.org/e4j/en/secondary/human-trafficking-and-migrant-smuggling.html# (Accessed: 17 October 2023)

Urbach, E. (2023) *Immigration. Petites phrases, gros préjugés.* Available at: https://www.humanite.fr/societe/immigration/immigration-petites-phrases-gros-prejuges-783697 (Accessed: 11 September 2023)

Van Dijk, T. A. (1989) ‘Mediating racism. The role of the media in the reproduction of racism’ in Wodak, R. (ed) *Language, power and ideology. Studies in political discourse*. Amsterdam: John Benjamins, pp. 199-226

Van Dijk, T. A. (1998) *Ideology: A multidisciplinary approach*. London: Sage

World Bank (2023) *Unemployment, total – Italy*. Available at: https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?locations=IT. (Accessed: 24 October 2023)

Wüster, E. (1968) *The machine tool: An interlingual dictionary of basic concepts*. London: Technical Press/Economic Commission for Europe of the United Nations

Wüster, E. (1991) *General introduction to terminology and terminological lexicography*. 3rd Edition. Bonn: Romanistischer Verlag

Zetter, R. (2007) 'More labels, fewer refugees: Remaking the refugee label in an era of globalization', *Journal of refugee studies*, 20(2), pp. 172-192

Zipf, G. K. (1949) *Human behavior and the principle of least effort*. Cambridge (Mass.): Addison-Wesley

Legal Sources

Allsports Ltd v UK office of fair trading (2003) United Kingdom competition appeal tribunal. Case 1021/1/03. Available at: https://www.catribunal.org.uk/cases/10211103-allsports-limited (Accessed: 10 October 2023)

Communication from the Commission to the European Parliament and the Council (2015) *A European Border and Coast Guard and effective management of Europe's external borders*, COM(2015) 673. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52015DC0673. (Accessed: 25 October 2023)

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions (2015) *A European Agenda on Migration,* COM(2015) 240. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015DC0240 (Accessed: 3 October 2023)

Consolidated version of the Treaty on the Functioning of the European Union (2010) *Official Journal* C 326/47, 30 March 2010, pp. 1-155

Consolidated version of Treaty establishing the European Community (2002) *Official Journal* C325, 24 December, pp.33-184

Convention relating to the status of refugees (1951) Treaty no. 2545. *United Nations Treaty Series*, 189, pp. 137-221. Available at: https://treaties.un.org/doc/Publication/UNTS/Volume%20189/volume-189-I-2545-English.pdf (Accessed: 3 October 2023)

Council conclusion 2005/C of 13 June 2005 on the official use of additional languages within the Council and possibly other Institutions and bodies of the European Union (2005) *Official Journal* C148, pp. 1-2

Council Decision (EU) 2015/1523 of 14 September 2015 establishing provisional measures in the area of international protection for the benefit of Italy and of Greece (2015) *Official Journal* L 239, pp. 146-156

Council Directive 2003/109/EC of 25 November 2003 concerning the status of third-country nationals who are long-term residents (2003) *Official Journal* L 016, 23/01/2004, pp. 44-53

Council of Europe Parliament Assembly (2006) *Resolution 1509 on Human Rights of Irregular Migrants.* Adopted on 27 June 2006

Council Directive 86/2003/EC of 22 September 2003 on the right to family reunification (2003) Official Journal L 251, pp.12-18

Council Directive 2009/50/EC of 25 May 2009 on the conditions of entry and residence of third-country nationals for the purposes of highly qualified employment (2009) *Official Journal* L 155, pp. 17-29

Council Regulation (EEC) 1958/1 determining the languages to be used by the European Economic Community (1958) *Official Journal* 017, pp. 385-386

Council Regulation (EU) No 517/2013 of 13 May 2013 (2013) *Official Journal,* L158, pp. 1-71

Decree no. 96-602 of 3 July 1996 on the enrichment of the French language (France)

Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC Text with EEA relevance (2014) *Official Journal* L 94, pp. 65-242

European Council (1999) *Tampere European Council 15 and 16 October 1999 Presidency Conclusions*. Available at: https://www.europarl.europa.eu/summits/tam_en.htm (Accessed: 3 October 2023)

European Council Conclusions (2019) *EUCO 29/19*. Available at: https://www.consilium.europa.eu/media/41768/12-euco-final-conclusions-en.pdf (Accessed: 17 October 2023)

European Parliament Resolution 2007/2145(INI) (2009) of 14 January on the situation of fundamental rights in the European Union 2004-2008. Available at: https://www.europarl.europa.eu/doceo/document/TA-6-2009-0019_EN.html. (Accessed: 14 September 2023)

Law no. 94-665 of 4 August 1994 relating to usage of the French language (France)

Immigration Act 1971*, c. 77*. Available at: https://www.legislation.gov.uk/ukpga/1971/77/contents. (Accessed: 24 October 2023)

International Convention on the Rights of all Migrant Workers and Members of Their Families (1990) *General Assembly resolution* 45/158. Available at: https://www.ohchr.org/en/instruments-mechanisms/instruments/international-convention-protection-rights-all-migrant-workers# (Accessed: 17 October 2023)

JJB Sports PLC v UK Office of Fair Trading (2004) United Kingdom Competition Appeal Tribunal. Case 1022/1/1/03. Available at: https://www.catribunal.org.uk/cases/10221103-jjb-sports-plc (Accessed: 10 October 2023)

Opinion of the Economic and Social Committee (1996) on the 'Proposal for a Council Directive on the right of third-country nationals to travel in the Community'. *Official Journal* C 153, pp. 38

Opinion of the European Economic and Social Committee (2016) Integration of refugees in the EU*. Official Journal* C 264, pp. 19-27

Proposal for a Council Act (1997) establishing the Convention on rules for the admission of third-country nationals to the member states*. Official Journal* C 337, pp. 9-19

Proposal for a Directive of the European Parliament and of the Council (2018) *Common standards and procedures in Member States for returning illegally staying third-country nationals (recast) A contribution from the European Commission to the Leaders' meeting in Salzburg on 19-20 September 2018.* COM/2018/634. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018PC0634. (Accessed: 25 October 2023)

Report from the Commission to the European Parliament and the Council (2016) *EU-Turkey Joint Action Plan - Third implementation report*, COM/2016/0144. Available at: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52016DC0144 (Accessed: 3 October 2023)

Resolution on the Communication from the Commission to the Council and the European Parliament on immigration and asylum policies (1995) *Official Journal C 362*, pp. 270

Srl CILFIT v Ministry of Health (1982) Case 283/81. ECLI:EU:C:1982:335. *European Court Reports* 1982 -03415

Treaty of Amsterdam amending the Treaty on European Union (1997) *Official Journal* C 340, 10 November, pp. 1-144

Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community (2007) *Official Journal* C 306, 13 December, pp. 1 -271

Treaty on European Union (1992) *Official Journal* C 191, 29 July, pp. 1-112

United Nations General Assembly (1967) *Protocol relating to the Status of Refugees*. Resolution 2198 (XXI) https://www.ohchr.org/sites/default/files/protocolrefugees.pdf (Accessed: 3 October 2023)

United Nations General Assembly (1975) *Measures to ensure the human rights of all migrant workers*. Resolution 3449. Available at: https://digitallibrary.un.org/record/189610?ln=en (Accessed: 17 October 2023)

Vienna Convention on the Law of Treaties (1969) *United Nations Treaty Series*, Vol. 1155. Available at: https://legal.un.org/ilc/texts/instruments/english/conventions/1_1_1969.pdf

Appendix 1.1: English to French *EUMigrLaw Corpus* 1992-1998

	ENG	FR	Frequency	Term Dist	Weighted term dist	IVI	Weighting	Mean weighted IVI
ASYLUM-SEEKER	Asylum-seeker	Demandeur d'asile	48	0	0			
	Asylum-seeker	OMITTED	1	1	1			
	Asylum seeker	Demandeur d'asile	21	0	0			
	Applicant for asylum	Demandeur d'asile	20	0	0			
	Applicant for asylum	Demande d'asile	1	0.5	0.5			
	Asylum applicant	Demandeur d'asile	1	0	0			
	Applicant	Demandeur d'asile	1	0.5	0.5			
Totals			93		2	IVI = 0.02151	X0.11369	0.00245
CROSS-BORDER WORKER	Transfrontier worker	Travailleur transfrontalier	2	0	0			
	Transfrontier worker	OMITTED	1	1	1			
			3		1	IVI = 0.33333	X0.00367	0.00122
DISPLACED PERSON	Displaced person	Personne deplacée	16	0	0			
	Displaced person	Déraciné	1	1	1			
	Displaced person	Personne expatrié	1	0.5	0.5			
Totals			18		1.5	IVI = 0.08333	x0.02200	0.00183
IRREGULAR MIGRANT	Illegal immigrant	Immigré en situation irregulière	8	1	8			
	Illegal immigrant	Personne en situation irregulière	5	1	5			
	Illegal immigrant	Immigrant illégal	4	0	0			
	Illegal immigrant	Immigré clandestin	3	0.5	1.5			
	Illegal immigrant	Immigré illégal	2	0	0			
	Illegal immigrant	Etranger entré illégalement	1	1	1			
	Illegal immigrant	entré de manière illicite	1	1	1			
	Illegal immigrant	sans papiers	1	1	1			
	Illegal immigrant	Immigrant entré illégalement	1	0.5	0.5			
	Illegal immigrant	Etranger en situation irregulière	1	1	1			
	Illegal immigrant	Etranger clandestin	1	1	1			
	Irregular immigrant	Immigré en situation irregulière	6	0.5	3			
	Illegal migrant	Illegal migrant	3	0	0			
Totals			37		23	IVI = 0.62162	x0.04523	0.02812

LONG-TERM RESIDENT	Long-term resident	(personne reconnue comme) durablement installée	16	1	16			
Totals			16		16	IVI = 1	x0.01956	0.01956
MIGRANT	Migrant	Migrant	45	0	0			
	Migrant	Immigrant	4	0.5	2			
	Migrant	Immigré	4	0.5	2			
	Migrant	Emigrant	1	0.5	0.5			
	Migrant	Refugié	1	1	1			
	Migrant	Immigration	1	1	1			
	Emigrant	Emigrant	1	0	0			
	Immigrant	Immigré	66	0	0			
	Immigrant	Immigrant	56	0	0			
	Immigrant	Emigrant	6	0.5	3			
	Immigrant	Migrant	4	0.5	2			
	Immigrant	Etranger	4	1	4			
	Immigrant	Emigré	1	0.5	0.5			
	Immigrant	Immigration	2	0.5	1			
	Immigrant	Ressortissant de pays tiers	1	1	1			
Totals			197		18	IVI = 0.09137	x0.24083	0.022
MIGRANT WOMAN	Migrant woman	Femme immigrée	1	0.5	0.5			
	Migrant woman	Femme migrante	2	0	0			
	Female immigrant	Femme migrante	1	0.5	0.5			
	Female migrant	Femme migrante	1	0	0			
Totals			5		1	IVI = 0.2	x0.00611	0.00122
MIGRANT WORKER	Migrant worker	Traveilleur migrant	19	0	0			
	Migrant worker	Travailleur clandestin	1	1	1			
	Migrant worker	Migrant	1	1	1			
	Immigrant worker	Travailleur immigré	1	0	0			
	Immigrant worker	Immigre	1	1	1			
	Immigrant worker	Travailleur immigrant	2	0	0			
	Immigrant worker	Main-d'oeuvre immigre	1	0.5	0.5			
	Foreign worker	Travailleur étranger	3	0	0			
	Foreign worker	Etranger	1	1	1			
	Foreign worker	Travailleur	1	1	1			
	Economic migrant	Migrant économique	3	0	0			

Totals			34		5.5	IVI = 0.16176	x0.04156	0.00672
REFUGEE	Refugee	Refugié	98	0	0			
	Refugee	OMITTED	1	1	1			
	Refugee	Personne fuyant	1	1	1			
Totals			100		2	IVI = 0.02	x0.12225	0.00245
RETURNEE	Returnee	Candidat au retour	2	1	2			
	Returnee	Celles qui retournent dans leur pays d'origine	1	1	1			
	Deportee	Personnes explusées	1	0.5	0.5			
Totals			4		3.5	IVI = 0.8755	x0.00367	0.00321
SEASONAL WORKER	Seasonal worker	Travailleur saisonnier	5	0	0			
Totals			5		0	IVI = 0		0
THIRD-COUNTRY NATIONAL	Third-country national	Ressortissant de pays tiers	125	0	0			
	Third-country national	Ressortissant d'un pays tiers	10	0	0			
	Third-country national	Personne de nationalite étrangère	1	0.5	0.5			
	Third-country national	Citoyen de pays tiers	2	0	0			
	Third-country national	Ressortissant d'Etats tiers	2	0	0			
	Third-country national	Citoyen étranger	1	0.5	0.5			
	Third-country national	Ressortissant non communautaire	1	0.5	0.5			
	Third-country national	ELLIPSIS	1	0.5	0.5			
	Alien	Etranger	9	0.5	4.5			
	Foreigner	Etranger	14	0	0			
	Foreigner	OMITTED	1	1	1			
	Citizen of third countries	Ressortissants de pays tiers	3	0	0			
	Citizen of third countries	Citoyen de pays tiers	1	0	0			
	Foreign national	Etranger	9	0.5	4.5			
	Foreign national	Ressortissant étranger	6	0	0			
	Foreign national	Personne étrangère	2	0.5	1			
	National of third countries	Ressortissant des pays tiers	10	0	0			
	National of third countries	Ressortissant de pays tiers	8	0	0			
Totals			206		13	IVI = 0.06311	x0.25183	0.01589
UNACCOMPANIED MINOR	Unaccompanied minor	Mineur non accompagné	5	0	0			
	Unaccompanied child	Enfant non accompagné	1	0	0			

Totals			6		0	IVI = 0		0
UNION CITIZEN	Citizen of the Union	Citoyen de l'Union	30	0	0			
	EU citizen	Citoyen de l'Union	7	0.5	3.5			
	EU citizen	Citoyen de l'UE	3	0	0			
	EU citizen	Ressortissant de l'Union	2	0.5	1			
	EU citizen	Ressortissant de l'Union européenne	1	0.5	0.5			
	EU citizen	Citoyen de l'Union européenne	1	0.5	0.5			
	EU citizen	Citoyen des états members	1	1	1			
	EU citizen	Ressortissant communautaire	1	1	1			
	EU citizen	Citoyen communautaire	1	1	1			
	European citizen	Citoyen européen	8	0	0			
	European citizen	Citoyen	3	1	3			
	European citizen	Citoyen de l'Europe	1	0	0			
	Union citizen	Citoyen de l'Union	6	0	0			
	National of member states	Ressortissant des états membres	3	0	0			
	National of member states	Ressortissant d'un état membre	3	0	0			
	National of member states	National des états membres	1	0	0			
Totals			72		11.5	IVI = 0.15972	x0.08802	0.01406
VICTIM OF TRAFFICKING IN HUMAN BEINGS	Victim of trafficking	Victime de la traite	18	0	0			
	Victim of trafficking	Victime	3	1	3			
	Victim of trafficking	Victime de la traite des êtres humains	1	0.5	0.5			
	Victim of trafficking in human beings	Victime de la traite des êtres humains	1	0	0			
Totals			23		3.5	IVI = 0.15217	x0.02812	0.00428

0.008200667	(Mean weighted IVI)

Appendix 1.2: French to Italian *EUMigrLaw Corpus* 1992-1998

	Fr	It	Frequency	Term Dist	Weighted term dist	IVI	Weighting	Weighted mean IVI
ASYLUM-SEEKER	Demandeur d'asile	Richiedente asilo	74	0	0			
	Demandeur d'asile	Richiedente l'asilo	8	0	0			
	Demandeur d'asile	persone che chiedono asilo	2	1	2			
	Demandeur d'asile	OMITTED	1	1	1			
	Demandeur d'asile	Richiedente	1	0.5	0.5			
	Demandeur d'asile	Coloro che presentano richiesta di asilo	1	1	1			
Totals			87		4.5	IVI = 0.05172	X0.10407	0.00538
CROSS-BORDER WORKER	Travailleur transfrontalier	Lavoratore transfrontaliero	2	0	0			
			2		0	IVI		0
DISPLACED PERSON	Personne déplacée	Sfollato	12	0.5	6			
	Personne déplacée	Profugo	2	0.5	1			
	Personne déplacée	Persona sfollata	1	0	0			
	Personne déplacée	Rifugiato	1	1	1			
	Personne expatriée	Sfollato	1	1	1			
	Déraciné	Sfollato	1	0.5	0.5			
Totals			18		9.5	IVI = 0.52778	X0.02153	0.01136
IRREGULAR MIGRANT	Personne en situation irrégulière	Persona in situazione irregolare	4	0	0			
	Personne en situation irrégulière	Persona in posizione irregolare	1	0.5	0.5			
	Immigrant illégal	Immigrato clandestino	2	0.5	1			
	Immigrant illégal	Immigrante clandestino	2	0.5	1			
	Immigré en situation irrégulière	Immigrato in situazione irregolare	18	0	0			
	Immigré clandestin	Immigrato clandestino	2	0	0			
	Immigré clandestin	Immigrazione clandestina	1	0.5	0.5			
	Sans-papier	Senza documenti	2	0	0			
	Sans-papier	Sans-papier	1	0	0			
	Sans-papier	Coloro che sono privi di documenti	1	1	1			

	Etranger en situation irrégulière	Straniero in situazione irregolare	3	0	0			
Totals			37		4	IVI = 0.10811	x0.04426	0.00478
LONG-TERM RESIDENT	(personne reconnue comme) durablement installée	(persone stabilitasi) a titolo duraturo	18	0.5	9			
			18		9	IVI = 0.5	X0.02153	0.01077
MIGRANT	Migrant	Migrante	41	0	0			
	Migrant	Immigrato	5	0.5	2			
	Migrant	Immigrante	1	0.5	0.5			
	Migrant	Emigrazione	1	1	1			
	Emigrant	Immigrato	4	1	4			
	Emigrant	Emigrante	1	0	0			
	Emigrant	Immigrante	1	1	1			
	Immigrant	Immigrato	35	0	0			
	Immigrant	Immigrante	20	0	0			
	Immigrant	Emigrato	2	1	2			
	Immigrant	Migrante	2	0.5	0.5			
	Immigrant	Cittadino di paesi terzi	1	1	1			
	Immigrant	Immigrazione	1	0.5	0.5			
	Immigré	Immigrato	61	0	0			
	Immigré	Migrante	2	0.5	1			
	Immigré	Immigrazione	1	0.5	0.5			
Totals			179		14	IVI = 0.07821	x0.21411	0.01675
MIGRANT WOMAN	Femme migrante	Donna migrante	5	0	0			
Totals			5		0	IVI = 0		0
MIGRANT WORKER	Travailleur migrant	Lavoratore migrante	21	0	0			
	Travailleur migrant	Migrante	2	0.5	1			
	Travailleur immigre	Lavoratore immigrate	1	0	0			
	Travailleur immigrant	Lavoratore immigrato	2	0	0			
	Travailleur étranger	Lavoratore straniero	3	0	0			
	Migrant économique	Cittadini di paesi terzi che cercano di entrare […] per motivi economici	1	1	1			
	Migrant économique	Altri che emigrano per motivi economici	1	1	1			
	Migrant économique	Emigrante per motivi economici	1	1	1			
Totals			32		4	IVI = 0.125	x0.03828	0.00479

REFUGEE	Refugié	Profugo	59	0.5	29.5			
	Refugié	Rifugiato	55	0	0			
	Refugié	Persona emigrante	1	1	1			
	Refugié	Profugo e rifugiato	1	0.5	0.5			
	Refugié	ELLIPSIS	3	0.5	1.5			
Totals			119		32.5	IVI = 0.27311	x0.14234	0.03887
RETURNEE	Candidat au retour	Persona che intende rientrare	2	0.5	1			
	Celles qui retournent	Coloro che fanno ritorno	1	0	0			
			3		1	IVI = 0.33	X0.0023 9	0.00079
SEASONAL WORKER	Travailleur saisonnier	Lavoratore stagionale	6	0	0			
Totals			6		0	IVI = 0		0
THIRD-COUNTRY NATIONAL	Ressortissant étranger	Cittadino straniero	4	0	0			
	Ressortissant étranger	Straniero	1	0.5	0.5			
	Citoyen du pays tiers	Cittadino dei paesi terzi	3	0	0			
	Ressortissant de pays tiers	Cittadino dei paesi terzi	177	0	0			
	Ressortissant de pays tiers	Cittadino di stati terzi	4	0.5	2			
	Ressortissant de pays tiers	Cittadino extracomunitario	1	0.5	0.5			
	Ressortissant de pays tiers	Cittadino di stati membri	1	1	1			
	Ressortissant de pays tiers	OMITTED	1	1	1			
	Ressortissant de pays tiers	ELLIPSIS	1	0.5	0.5			
	Ressortissant d'un pays tiers	Cittadino proveniente da un paese terzo	1	0.5	0.5			
	Etranger	Straniero	34	0	0			
	Etranger	OMITTED	1	1	1			
	Etranger	Cittadino straniero	1	0.5	0.5			
Totals			232		7.5	IVI = 0.03233	x0.27751	0.00897
UNACCOMPANIED MINOR	Mineur non accompagné	Minorenne non accompagnato	3	0	0			
	Mineur non accompagné	Minore non accompagnato	3	0	0			
	Enfant non accompagné	Minore non accompagnato	1	0.5	0.5			
Totals			7		0.5	IVI = 0.07143	X0.0083 7	0.00006
UNION CITIZEN	Ressortissant communautaire	Cittadino comunitario	4	0	0			

	Ressortissant communautaire	Cittadino degli stati membri	1	0.5	0.5			
	Citoyen européen	Cittadino europeo	6	0	0			
	Citoyen européen	Cittadino dell'Unione	1	0.5	0.5			
	Citoyen de l'Union	Cittadino dell'Unione	34	0	0			
	Citoyen de l'Union	Cittadino dell'Unione europea	3	0.5	1.5			
	Citoyen de l'Union européenne	Cittadino dell'Unione europea	4	0	0			
	Citoyen de l'Union européenne	Cittadino dell'Unione	1	0.5	0.5			
	Citoyen de l'Union européenne	Cittadino dell'UE	1	0.5	0.5			
	Citoyen de l'UE	Cittadino dell'Unione europea	2	0.5	1			
	Citoyen de l'UE	Cittadino comunitario	1	0.5	0.5			
	Citoyen des Etats membres	Cittadino degli stati membri	4	0	0			
	Citoyen communautaire	Cittadino comunitario	4	0	0			
Totals			66		6	IVI = 0.09091	x0.0789 5	0.00718
VICTIM OF TRAFFICKING IN HUMAN BEINGS	Victime de la traite	Vittima del traffico	14	1	14			
	Victime de la traite	Vittima della tratta	6	0	0			
	Victime de la traite	OMITTED	1	1	1			
	Victime de la traite	Vittima	2	0.5	1			
	Victime de la traite des êtres humains	Vittima di tratta	1	0.5	0.5			
	Victime de la traite des êtres humains	Vittima della tratta degli esseri umani	1	0	0			
			26		16.5	IVI = 0.63462	x0.0311 0	0.01974
								0.008 66533 3

Mean weighted IVI

Appendix 1.3: Italian to English *EUMigrLaw Corpus* 1992-1998

	IT	ENG	Frequency	Term Dist	Weighted term dist	IVI	Weighting	Weighted mean IVI
ASYLUM-SEEKER	Richiedente asilo	Asylum-seeker	36	0	0			
	Richiedente asilo	Applicant for asylum	19	0	0			
	Richiedente asilo	Asylum seeker	16	0	0			
	Richiedente asilo	Asylum application	1	0.5	0.5			
	Richiedente asilo	Asylum applicant	1	0	0			
	Richiedente asilo	Applicant (persons) seeking asylum	1	0.5	0.5			
	Richiedente asilo		1	0.5	0.5			
	Richiedente l'asilo	Asylum-seeker	4	0	0			
	Richiedente l'asilo	persons seeking asylum	1	0.5	0.5			
Totals			80		2	IVI = 0.025	x0.09744	0.00244
CROSS-BORDER WORKER	Lavoratore transfrontaliero	Transfrontier worker	3	0	0			
			3		0	IVI = 0		0
DISPLACED PERSON	Sfollato	Displaced person	14	0.5	7			
	Persona sfollata	Displaced person	1	0				
Totals			15		7	IVI = 0.46667	x0.01827	0.00853
IRREGULAR MIGRANT	Persona in situazione irregolare	Illegal immigrant	4	1	4			
	Immigrato in situazione irregolare	Illegal immigrant	8	1	8			
	Immigrato in situazione irregolare	Irregular immigrant	6	0.5	3			
	Immigrato in situazione irregolare	Immigrant with irregular status	2	0.5	1			
	Immigrato in situazione irregolare	Immigrant whose status is irregular	1	0.5	0.5			
	Immigrato in situazione irregolare	Immigrant whose situation is not in order	1	1	1			
	Immigrato clandestino	Illegal immigrant	4	0.5	2			
	Immigrato clandestino	Illegal migrant	1	1	1			
	Clandestino	Illegal immigrant	2	1	2			
	Clandestino	Illegal workers	2	1	2			

	Clandestino	those who are not legally resident	1	1	1			
	Clandestino	Human being	1	1	1			
	Clandestino	Illegal immigration	1	1	1			
	Clandestino	Illegal migrant	1	1	1			
	Clandestino	Clandestine migrant	1	0.5	0.5			
	Sans papiers	Illegal immigrant	2	1	2			
	Sans-papiers	Sans papiers	1	0	0			
Totals			39		31	IVI = 0.79487	x0.0475	0.03776
LONG-TERM RESIDENT	(Persona stabilitasi) a titolo duraturo	Long-term resident	15	0.5	7.5			
Totals			15		7.5	IVI = 0.5	x0.01827	0.00914
MIGRANT	Migrante	Migrant	38	0	0			
	Migrante	Immigrant	5	0.5	2.5			
	Migrante	Refugee	1	1	1			
	Migrante	Immigration	1	1	1			
	Emigrante	Migrant	4	0.5	2			
	Emigrante	Emigrant	1	0	0			
	Immigrato	Immigrant	90	0	0			
	Immigrato	Migrant	7	0.5	3.5			
	Immigrante	Immigrant	20	0	0			
	Immigrante	Immigration	3	0.5	1.5			
	Immigrante	Migrant	1	0.5	0.5			
Totals			171		12	IVI = 0.07018	x0.20828	0.01462
MIGRANT WOMAN	Donna migrante	Migrant women	3	0	0			
	Donna migrante	Female immigrant	1	0	0			
	Donna migrante	Female migrant	1	0	0			
	Donna migrante	Immigrant's wife	1	1	1			
Totals			6		1	IVI = 0.16667	x0.00731	0.00122
MIGRANT WORKER	Lavoratore migrante	Migrant worker	23	0	0			
	Lavoratore migrante	Migrant	5	0.5	2.5			
	Lavoratore migrante	Immigrant	1	1	1			
	Lavoratore straniero	Foreign worker	3	0	0			
	Lavoratore immigrato	Immigrant workers	3	0	0			
	Cittadini di paesi terzi che cercano di entrare […] per motivi economici	Economic migrant	1	1	1			

	Altri che emigrano per motivi economici	Economic migrant	1	1	1			
	Emigranti per motivi economici	Economic migrant	1	1	1			
Totals			38		6.5	IVI = 0.17105	x0.04629	0.00792
REFUGEE	Rifugiato	Refugee	53	0	0			
	Rifugiato	Displaced person	1	0.5	0.5			
	Rifugiato	ELLIPSIS	4	0.5	2			
	Profugo	Refugee	63	0.5	21.5			
	Profugo	Displaced person	1	0.5	0.5			
	Profugo	ELLIPSIS	3	0.5	1.5			
Totals			125		26	IVI = 0.208	x0.1522 5	0.03167
RETURNEE	Persona che intende rientrare	Returnee	2	0.5	1			
			2		1	IVI = 0.5	X0.0024 4	0.00122
SEASONAL WORKER	Lavoratore stagionale	Seasonal worker	5	0	0			
	Lavoratore stagionale	Seasonal work	1	0.5	0.5			
Totals			6		0.5	IVI = 0.08333	x0.0073 1	0.00061
THIRD-COUNTRY NATIONAL	Cittadino di paesi terzi	Third-country national	127	0	0			
	Cittadino di paesi terzi	National of third countries	23	0	0			
	Cittadino di paesi terzi	Third country national	12	0	0			
	Cittadino di paesi terzi	Citizen of third countries	3	0	0			
	Cittadino di paesi terzi	National of non-member countries	3	0.5	1.5			
	Cittadino di paesi terzi	Immigrant	1	1	1			
	Cittadino di paesi terzi	Immigrant from non-member countries	1	1	1			
	Cittadino di paesi terzi	Citizen of non-member countries	1	0.5	0.5			
	Cittadino straniero	Third-country national	2	0.5	1			
	Cittadino straniero	Foreign national	2	0	0			
	Straniero	Foreign national	12	0.5	6			
	Straniero	Foreigner	9	0	0			
	Straniero	Alien	7	0.5	3.5			
	Straniero	Immigrant	2	0.5	1			
	Straniero	Third-country national	1	0.5	0.5			

	Straniero	Non-national	1	0.5	0.5			
	Cittadino extracomunitario	Third-country national	1	0.5	0.5			
	Cittadino extracomunitario	non-EU citizen	1	0.5	0.5			
	Cittadino extracomunitario	Non-community residents	1	0.5	0.5			
Totals			207		18	IVI = 0.08696	x0.2521 3	0.02193
UNACCOMPANIED MINOR	Minore non accompagnato	Unaccompanied minor	2	0	0			
	Minore non accompagnato	Unaccompanied child	2	0.5	1			
	Minore non accompagnato	Refugee child	1	1	1			
	Minorenne non accompagnato	Unaccompanied minor	3	0	0			
Totals			8		2	IVI = 0.25	x0.00974	0.00237
UNION CITIZEN	Cittadino dell'Unione	Citizen of the Union	26	0	0			
	Cittadino dell'Unione	EU citizen	7	0.5	3.5			
	Cittadino dell'Unione	Union citizen	5	0	0			
	Cittadino dell'Unione	the public	1	1	1			
	Cittadino dell'Unione	Resident of the Union	1	0.5	0.5			
	Cittadino dell'Unione	European citizen	1	0.5	0.5			
	Cittadino degli stati membri	Nationals of member states	10	0	0			
	Cittadino degli stati membri	EU national	1	0.5	0.5			
	Cittadino degli stati membri	Citizen of a member state	1	0	0			
	Cittadino degli stati membri	EU citizen	1	0.5	0.5			
	Cittadino degli stati membri	Member State national	1	0	0			
	Cittadino degli stati membri	resident of a member state	1	0.5	0.5			
	Cittadino comunitario	EU citizen	3	0.5	1.5			
	Cittadino comunitario	Community national	2	0	0			
	Cittadino comunitario	EC national	2	0.5	1			
	Cittadino comunitario	Community citizen	1	0	0			
	Cittadino comunitario	EU national	1	0.5	0.5			
	Cittadino comunitario	non-EU citizen	1	1	1			
	Cittadino europeo	European citizen	5	0	0			
	Cittadino europeo	Citizen of Europe	2	0	0			
	Cittadino dell'Unione europea	EU citizen	8	0.5	4			
	Cittadino dell'Unione europea	European Union citizen	1	0	0			
	Cittadino dell'Unione europea	Citizen of the European Union	1	0	0			
	Cittadino dell'Unione europea	EU national	1	0.5	0.5			
	Cittadino dell'Unione europea	Citizen of the Union	1	0.5	0.5			

						IVI = 0.18824	x0.1035 3	0.01949
Totals			85		16			
VICTIM OF TRAFFICKING IN HUMAN BEINGS	Vittima del traffico	Victim of trafficking	14	1	14			
	Vittima della tratta	Victim of trafficking	6	0	0			
	Vittima della tratta	Victim	1	0.5	0.5			
			21		14.5	IVI = 0.69048	x0.0255 8	0.01766
								0.011772

Mean weighted IVI

Appendix 1.4: English to French *EUMigrLaw Corpus* 2015-2018

	ENG	FR	Frequency	Term Dist	Weighted Term Dist	IVI	Weighting	Weighted mean IVI
ASYLUM-SEEKER	Asylum seeker	Demandeur d'asile	52	0	0			
	Asylum seeker	Demandeuse d'asile	1	0.5	0.5			
	Asylum-seeker	Demande d'asile	1	0.5	0.5			
	Asylum-seeker	Demandeur d'asile	16	0	0			
	Asylum applicant	Demandeur d'asile	10	0	0			
	Applicant for international protection	Demandeur de protection internationale	9	0	0			
	Applicant for international protection	Demandeur d'une protection internationale	8	0	0			
	Applicant for international protection	Personne demandant la protection internationale	2	0.5	1			
	Applicant for international protection	Candidats à une protection internationale	2	0.5	1			
Totals			101		3	IVI = 0.02970	x0.05390	0.0016
CROSS-BORDER WORKER	Cross-border worker	Travailleur frontalier	8	0	0			
	Cross-border/frontier worker	Travailleur frontalier	1	0.5	0.5			
Totals			9		0.5	IVI = 0.05556	x0.0048	0.00027
DISPLACED PERSON	Displaced person	Personne deplacée	20	0	0			
	Displaced person	Déplacement de population	1	0.5	0.5			
	The displaced	Personne deplacée	1	0.5	0.5			
Totals			22		1	IVI = 0.04545	x0.01174	0.00053
IRREGULAR MIGRANT	Irregular migrant	Migrant en situation irregulière	37	0.5	18.5			
	Irregular migrant	Migrant irregulier	2	0	0			
	Irregular migrant	Migrant clandestin	2	0.5	1			
	Irregular migrant	ELLIPSIS	1	0.5	0.5			
	Irregular migrant	Immigrant clandestin	1	1	1			
	Irregular migrant	Migrant en séjour irregulier	1	0.5	0.5			
	Illegal immigrant	Immigration illégale	3	0.5	1.5			
	Illegal immigrant	Immigré illégal	2	0	0			
	Illegal immigrant	Immigrant illégal	1	0	0			

	Illegal immigrant	Migrant	1	1	1			
	Illegal immigrant	Immigré en situation illegale	1	0.5	0.5			
	Unauthorised migrant	Migrant en situation irregulière	2	1	2			
	Persons residing without authorisation	Personne en séjour irregulier	2	0.5	1			
Totals			56		27.5	IVI = 0.49107	x0.02988	0.01467
LONG-TERM RESIDENT	Long-term resident	Résident de longue durée	3	0	0			
Totals			3		0	IVI = 0		0
MIGRANT	Migrant	Migrant	402	0	0			
	Migrant	OMITTED	3	1	3			
	Migrant	Personne	2	1	2			
	Migrant	Migration	2	0.5	1			
	Migrant	Immigrant	1	0.5	0.5			
	Migrant	Migratoire	1	0.5	0.5			
	Migrant	Personne immigrée	1	0.5	0.5			
	Migrant	Immigration	1	1	1			
	Immigrant	Immigrant	5	0	0			
	Immigrant	Immigré	5	0	0			
	Immigrant	Migrant	5	0.5	2.5			
	Immigrant	Immigration	3	0.5	1.5			
	Immigrant	Ressortissant	1	1	1			
	Newcomer	Nouvel arrivant	6	0	0			
	Newcomer	Nouvel arrivé	1	0	0			
	Newcomer	Nouveau venu	1	0	0			
Totals			440		13.5	IVI = 0.03068	x0.23479	0.00721
MIGRANT CHILD (IOM)	Child in migration	Enfant migrant	4	0.5	2			
	Migrant child	Enfant migrant	4	0	0			
Totals			8		2	IVI = 0.25000	x0.00427	0.00107
MIGRANT WORKER	Migrant worker	Travailleur migrant	4	0	0			
	Foreign worker	Travailleur étranger	4	0	0			
	Economic migrant	Migrant économique	7	0	0			
	Economic migrant	Immigrant économique	1	0.5	0.5			
Totals			16		0.5	IVI = 0.03125	x0.00854	0.00027
MOBILE WORKER	Mobile worker	Travailleur mobile	13	0	0			
	Mobile worker	Personne mobile	1	0.5	0.5			

Totals			14		0.5	IVI = 0.03571	x0.00747	0.00027
OFFSHORE WORKER	Offshore worker	Travailleur offshore	4	0	0			
Totals			4		0	IVI = 0		0
POSTED WORKER	Posted worker	Travailleur détaché	9	0	0			
Totals			9		0	IVI = 0		0
REFUGEE	Refugee	Refugié	355	0	0			
	Refugee	ELLIPSIS	4	0.5	2			
	Refugee	OMITTED	2	1	2			
	Beneficiary of internation protection	Bénéficiaire d'une protection internationale	14	0	0			
	Beneficiary of internation protection	Bénéficier d'une protection internationale	4	0.5	2			
Totals			379		6	IVI = 0.01583	x0.20224	0.0032
RELOCATED PERSON	Relocated person	Personne relocalisée	6	0	0			
	Relocated person	Relocalisation	1	0.5	0.5			
Totals			7		0.5	IVI = 0.07143	x0.00374	0.00027
RESETTLED PERSON	Person resettled	Personne reinstallée	6	0.5	3			
	Resettled person	Personne reinstallée	1	0	0			
			7		3	IVI = 0.42857	x0.00374	0.0016
RETURNEE	Returnee	Personne soumise à un retour	17	0.5	8.5			
	Returnee	Personne ayant fait l'objet d'une mesure de retour	3	0.5	1.5			
	Returnee	Personne rapatriée	2	0.5	1			
	Returnee	Rapatrié	1	0.5	0.5			
	Returnee	Personne faisant l'objet d'une mesure de retour	1	0.5	0.5			
	Returnee	ELLIPSIS	1	0.5	0.5			
	Returnee	Personne	1	1	1			
	Returnee	Personne faisant l'objet d'une décision de retour	1	0.5	0.5			
Totals			27		14	IVI = 0.51852	x0.01441	0.00747
SEASONAL WORKER	Seasonal worker	Travailleur saisonnier	4	0	0			
Totals			4		0	IVI = 0		0

STATELESS PERSON	Stateless person	Apatride	32	0.5	16			
	Stateless person	Personne apatride	2	0	0			
Totals			34		16	IVI = 0.47059	x0.01814	0.00854
THIRD-COUNTRY NATIONAL	Third-country national	Ressortissant de pays tiers	480	0	0			
	Third-country national	Ressortissant d'un pays tiers	48	0	0			
	Third-country national	Ressortissant du pays tiers	8	0	0			
	Thirdcountry national	Ressortissant de pays tiers	5	0	0			
	National of third country	Ressortissant de pays tiers	10	0	0			
	National of third country	Ressortissant des pays tiers	4	0	0			
	National of third country	Ressortissant provenant de pays tiers	1	0.5	0.5			
Totals			556		0.5	IVI = 0.00090	x0.29669	0.00027
UNACCOMPANIED MINOR	Unaccompanied minor	Mineur non accompagné	48	0	0			
	Unaccompanied child	Enfant non accompagné	3	0	0			
	Refugee child	Enfant refugié	6	0	0			
	Refugee child	Enfant de refugiés	1	0.5	0.5			
Totals			58		0.5	IVI = 0.00862	x0.30950	0.00267
UNION CITIZEN	Union citizen	Citoyen de l'Union	13	0	0			
	Citizen of the Union	Citoyen de l'Union	12	0	0			
	EU citizen	Citoyen de l'Union	9	0.5	4.5			
	EU citizen	Citoyen de l'UE	3	0	0			
	EU citizen	Citoyen de l'Union Européenne	2	0.5	1			
	European citizen	Citoyen européen	4	0	0			
	European citizen	Citoyen de l'Union	1	0.5	0.5			
	European citizen	Citoyen de l'Union européenne	1	0.5	0.5			
	European	Citoyen européen	2	0.5	1			
	European	Européen	2	0	0			
Totals			49		7.5	IVI = 0.15306	x0.02615	0.004
VICTIM OF TRAFFICKING IN HUMAN BEINGS	Victim of trafficking in human beings	Victime de la traite des êtres humains	21	0	0			
	Victim of trafficking in human beings	Victime de la traite	4	0.5	2			
	Victim of trafficking	Victime de la traite	15	0	0			
	Victim of trafficking	Victime de la traite des êtres humains	4	0.5	2			
	Victim of trafficking	Victime des trafiquants d'êtres humains	1	0.5	0.5			

	Human trafficking victim	Victime de la traite des êtres humains	2	0	0				
	Victim of human trafficking	Victime de la traite des êtres humains	6	0	0				
	Victim of human trafficking	Victime de la traite	1	0.5	0.5				
Totals			54		5	IVI = 0.09259	x0.028 82	0.0026 7	
VULNERABLE PERSON	Vulnerable person	Personne vulnérable	16	0	0				
	Vulnerable person	Personne en situation vulnérable	1	0.5	0.5				
			17		0.5	IVI = 0.02941	x0.009 07	0.0002 7	
								0.0025 83636	Mean weighted IVI

Appendix 1.5: French to Italian *EUMigrLaw Corpus* 2015-2018

	Fr	It	Frequency	Term Dist	Weighted term dist	IVI	Weighting	Weighted mean IVI
ASYLUM-SEEKER	Demandeur d'asile	Richiedente asilo	64	0	0			
	Demandeur d'asile	Domanda di asilo	3	0.5	1.5			
	Demandeur d'asile	Coloro che cercano asilo	1	0.5	0.5			
	Demandeur de protection internationale	Richiedente protezione internazionale	9	0	0			
	Demandeur de protection internationale	Richiedente	1	0.5	0.5			
	Demandeur d'une protection internationale	Richiedente protezione internazionale	8	0	0			
Totals			86		2.5	IVI = 0.02907	x0.04567	0.00133
CROSS-BORDER WORKER	Travailleur frontalier	Lavoratore frontaliere	6	0	0			
	Travailleur frontalier	Lavoratore transfrontaliere	2	0.5	1			
	Travailleur frontalier	Lavoratore	1	0.5	0.5			
Totals			9		1.5	IVI = 0.16667	x0.00478	0.0008
DISPLACED PERSON	Personne deplacée	Sfollato	19	0.5	9.5			
	Personne deplacée	Persona	2	1	2			
	Personne deplacée	Persona sfollata	2	0	0			
	Personne deplacée	Sfollamento	1	1	1			
	Personne deplacée	Persona che si sposta	1	0.5	0.5			
	Personne deplacée	Vittima di sfollamento	1	1	1			
	Personne deplacée	Persona che e stata costratte a sfollare	1	1	1			
Totals			27		15	IVI = 0.55556	x0.01434	0.008
IRREGULAR MIGRANT	Migrant en situation irrégulière	Migrante irregolare	34	0.5	17			
	Migrant en situation irrégulière	Migrante clandestino	2	1	2			
	Migrant en situation irrégulière	Migrazione irregolare	1	1	1			
	Migrant en situation irrégulière	Migrante in posizione irregolare	1	0.5	0.5			
	Personne en séjour irrégulier	OMITTED	7	1	7			
	Personne en séjour irrégulier	Persona in soggiorno irregolare	2	0	0			

	Migrant irrégulier	Migrante irregolare	3	0	0			
Totals			50		27.5	IVI = 0.55000	x0.02655	0.0146
LONG-TERM RESIDENT	Resident de longue durée	Soggiornante di lungo periodo	3	0	0			
Totals			3		0	IVI = 0		0
MIGRANT	Migrant	Migrante	397	0	0			
	Migrant	OMITTED	12	1	12			
	Migrant	Immigrato	5	0.5	2.5			
	Migrant	Arrivo	4	0.5	2			
	Migrant	Partenza	3	0.5	1.5			
	Migrant	Migrazione	3	0.5	1.5			
	Migrant	Essere umano	2	1	2			
	Migrant	Attraversamento	2	0.5	1			
	Migrant	Clandestino	1	1	1			
	Migrant	Persona migrante	1	0.5	0.5			
	Migrant	Quanti migrano	1	0.5	0.5			
	Migrant	Richiendente asilo	1	1	1			
	Migrant	Persona	1	1	1			
	Migrant	Migratorio	1	0.5	0.5			
	Personne migrante	Persone migrante	4	0	0			
	Personne migrante	Migrante	1	0.5	0.5			
	Immigrant	Immigrato	6	0	0			
	Immigrant	Immigrazione	1	0.5	0.5			
	Immigré	Immigrato	3	0	0			
	Immigré	Clandestino	2	1	2			
	Immigré	Immigrazione	1	0.5	0.5			
	Nouvel arrivant	Nuovo arrivato	6	0	0			
	Nouvel arrivé	Nuovo arrivato	1	0	0			
	Nouveau venu	Nuovo arrivato	1	0	0			
Totals			460		30.5	IVI = 0.06630	x0.24429	0.0162
MIGRANT CHILD	Enfant migrant	Minore migrante	7	0	0			
Totals			7		0	IVI = 0		0
MIGRANT WORKER	Travailleur migrant	Lavoratore migrante	5	0	0			
	Travailleur étranger	Lavoratore straniero	4	0	0			
	Migrant économique	Migrante economico	6	0	0			
	Migrant économique	Migrante per motivi economici	1	0.5	0.5			
	Immigrant économique	Migrante economico	1	0.5	0.5			
Totals			17		1	IVI = 0.05882	x0.00903	0.00053

MOBILE WORKER	Travailleur mobile	Lavoratore mobile	13	0	0			
Totals			13		0	IVI = 0		0
OFFSHORE WORKER	Travailleur offshore	Lavoratore off-shore	3	0	0			
	Travailleur offshore	Lavoratore offshore	1	0	0			
Totals			4		0	IVI = 0		0
POSTED WORKER	Travailleur détaché	Lavoratore distaccato	8	0	0			
Totals			8		0	IVI = 0		0
REFUGEE	Refugié	Rifugiato	342	0	0			
	Refugié	Profugo	17	0.5	8.5			
	Refugié	Domanda di rifugio	2	0.5	1			
	Refugié	Migrante	2	1	2			
	Refugié	Persona	1	1	1			
	Refugié	OMITTED	1	1	1			
	Refugié	Cittadino	1	1	1			
	Bénéficiaire d'une protection internationale	Beneficiario di protezione internazionale	14	0	0			
	Bénéficiaire d'une protection internationale	Protezione internazionale	1	0.5	0.5			
Totals			381		15	IVI = 0.03937	x0.20234	0.00797
RELOCATED PERSON	Personne relocalisée	Persona trasferita	2	0.5	1			
	Personne relocalisée	Persona	1	1	1			
	Personne relocalisée	Ricollocato	1	0.5	0.5			
	Personne relocalisée	Persona ricollocata	3	0	0			
Totals			7		2.5	IVI = 0.35714	x0.00372	0.00133
RESETTLED PERSON	Personne reinstallée	Persona reinsediata	8	0	0			
	Personne reinstallée	Reinsediamento	2	1	2			
Totals			10		2	IVI = 0.20000	x0.00531	0.00106
RETURNEE	Personne soumise à un retour	Rimpatriato	16	0.5	8			
	Personne soumise à un retour	Rimpatriando	1	0.5	0.5			
	Rapatrié	Rimpatriato	1	0	0			
	Personne faisant l'objet d'une mesure de retour	Rimpatriato	1	0.5	0.5			
	Personne faisant l'objet d'une decision de retour	Rimpatriando	1	0.5	0.5			
	Personne rapatriée	Rimpatriato	1	0.5	0.5			

Totals			21		10	IVI = 0.47619	x0.01115	0.00531
SEASONAL WORKER	Travailleur saisonnier	Lavoratore stagionale	4	0	0			
			4		0	IVI = 0		0
STATELESS PERSON	Apatride	Apolide	33	0	0			
			33		0	IVI = 0		0
THIRD-COUNTRY NATIONAL	Ressortissant de pays tiers	Cittadino di paesi terzi	574	0	0			
	Ressortissant de pays tiers	Immigrato proveniente da paese terzo	1	0.5	0.5			
	Ressortissant de pays tiers	Cittadino non avente la cittadinanza di uno stato membro	1	0.5	0.5			
Totals			576		1	IVI = 0.00174	x0.30589	0.00053
UNACCOMPANIED MINOR	Mineur non accompagné	Minore non accompagnato	42	0	0			
	Enfant non accompagné	Minore non accompagnato	3	0.5	1.5			
	Enfant refugié	Minore rifugiato	6	0	0			
	Enfant refugié	Bambino rifugiato	1	0	0			
Totals			52		1.5	IVI = 0.02885	x0.02762	0.00008
UNION CITIZEN	Citoyen de l'Union	Cittadino dell'Unione	16	0	0			
	Citoyen de l'Union	Cittadino dell'UE	6	0.5	3			
	Citoyen de l'Union	Cittadino dell'Unione europea	4	0.5	2			
	Citoyen de l'Union	Cittadino UE	1	0.5	0.5			
	Citoyen de l'Union	Cittadino europeo	1	0.5	0.5			
	Citoyen européen	Cittadino europeo	6	0	0			
	Citoyen européen	Cittadinanza europea	1	0	0			
	Citoyen de l'UE	Cittadino dell'UE	2	0	0			
	Citoyen de l'UE	Cittadino UE	1	0.5	0.5			
	Citoyen de l'Union européenne	Cittadino europeo	1	0.5	0.5			
	Citoyen de l'Union européenne	Cittadino dell'UE	1	0.5	0.5			
	Européen	Cittadino europeo	2	0.5	1			
Totals			42		8.5	IVI = 0.20238	x0.02230	0.00451
VICTIM OF TRAFFICKING IN HUMAN BEINGS	Victime de la traite des êtres humains	Vittima della tratta di esseri umani	25	0	0			
	Victime de la traite des êtres humains	Vittima della tratta	4	0.5	2			
	Victime de la traite	Vittima della tratta	19	0	0			
	Victime de la traite	Vittima della tratte di esseri umani	8	0.5	4			

	Victime de la traite	Traffico di persone	1	1	1			
						IVI = 0.12	x0. 030	0.00
Totals			57		7	281	27	372
VULNERABLE PERSON	Personne vulnérable	Persona vulnerabile	15	0	0			
	Personne vulnérable	Soggetto vulnerabile	1	0.5	0.5			
						IVI = 0.03	x0. 008	0.00
Totals			16		0.5	125	5	027

	Weighted
0.00	mean
3043	IVI
636	

Appendix 1.6: Italian to English *EUMigrLaw Corpus* 2015-2018

	IT	ENG	Frequency	Term dist	Weighted term dist	IVI	Weighting	Weighted mean IVI
ASYLUM-SEEKER	Richiedente asilo	Asylum seeker	48	0	0			
	Richiedente asilo	Asylum-seeker	13	0	0			
	Richiedente asilo	Asylum applicant	8	0	0			
	Richiedente asilo	Migrant	1	1	1			
	Richiedente asilo	Asylum application	1	0.5	0.5			
	Richiedente asilo	OMITTED	1	1	1			
	Richiedente asilo	Asylum flow	1	1	1			
	Richiedente protezione internazionale	Applicant for international protection	21	0	0			
	Richiedente protezione internazionale	Persons applied for international protection	1	0.5	0.5			
	Chi necessita di protezione internazionale	Those in need of international protection	2	0	0			
	Chi necessita di protezione internazionale	People in need of international protection	1	0	0			
Totals			98		4	IVI = 0.04082	x0.05224	0.00213
CROSS-BORDER WORKER	Lavoratore frontaliero	Cross-border worker	5	0	0			
	Lavoratore frontaliero	Cross-border/frontier worker	1	0.5	0.5			
	Lavoratore transfrontaliero	Cross-border worker	2	0	0			
			8		0.5	IVI = 0.0625	x0.00426	0.00027
DISPLACED PERSON	Sfollato	Displaced person	20	0.5	10			
	Sfollato	The displaced	1	0	0			
	Sfollato	Displacement	1	0.5	0.5			
Totals			22		10.5	IVI = 0.47727	x0.01173	0.0056
IRREGULAR MIGRANT	Migrante irregolare	Irregular migrant	40	0	0			
	Clandestino	Illegal immigrant	2	1	2			
	Clandestino	Immigrant	1	1	1			
	Clandestino	OMITTED	1	1	1			
	Migrante clandestino	Unauthorised migrant	2	0.5	1			
	Migrante in posizione irregolare	Irregular migrant	1	0.5	0.5			
	Persona in soggiorno irregolare	Persons residing without authorisation	2	0.5	1			

Totals			49		6.5	IVI = 0.13265	x0.02612	0.00346
LONG-TERM RESIDENT	Soggiornante di lungo periodo	Long-term resident	2	0	0			
	Soggiornante di lungo periodo	Long-term residence	1	0.5	0.5			
			3		0.5	IVI = 0.16667	x0.0016	0.00027
MIGRANT	Migrante	Migrant	383	0	0			
	Migrante	OMITTED	8	1	8			
	Migrante	Arrivals	3	0.5	1.5			
	Migrante	Refugee	2	1	2			
	Migrante	Migration	2	0.5	1			
	Migrante	Return	1	0.5	0.5			
	Migrante	People	1	1	1			
	Persona migrante	Migrant	17	0.5	8.5			
	Immigrato	Immigrant	11	0	0			
	Immigrato	Migrant	5	0.5	2.5			
	Nuovo arrivato	Newcomer	8	0	0			
Totals			441		25	IVI = 0.05669	x0.23507	0.01333
MIGRANT CHILD	Minore migrante	Child in migration	5	0.5	2.5			
	Minore migrante	Migrant child	2	0	0			
Totals			7		2.5	IVI = 0.35714	x0.00373	0.00133
MIGRANT WORKER	Lavoratore migrante	Migrant worker	5	0	0			
	Lavoratore straniero	Foreign worker	4	0	0			
	Migrante economico	Economic migrant	7	0	0			
	Migrante per motivi economici	Economic migrant	1	0.5	0.5			
Totals			17		0.5	IVI = 0.02941	x0.00906	0.00027
MOBILE WORKER	Lavoratore mobile	Mobile worker	14	0	0			
Totals			14		0	IVI = 0		0
OFFSHORE WORKER	Lavoratore off-shore	Offshore worker	3	0	0			
	Lavoratore offshore	Offshore worker	1	0	0			
Totals			4		0	IVI = 0		0
POSTED WORKER	Lavoratore distaccato	Posted worker	9	0	0			
Totals			9		0	IVI = 0		0

REFUGEE	Rifugiato	Refugee	339	0	0			
	Rifugiato	OMITTED	2	1	2			
	Profugo	Refugee	23	0.5	11.5			
	Beneficiario di protezione internazionale	Beneficiary of international protection	19	0	0			
Totals			383		13.5	IVI = 0.03525	x0.20416	0.00727
RELOCATED PERSON	Persona ricollocata	Relocated people	2	0	0			
	Persona ricollocata	Relocated person	1	0	0			
	Persona ricollocata	People have been relocated	1	0.5	0.5			
	Persona trasferita	Relocated person	2	0.5	1			
	Ricollocato	Relocated person	1	0.5	0.5			
Totals			7		2	IVI = 0.28571	x0.00373	0.00107
RESETTLED PERSON	Persona reinsediata	Person resettled	7	0.5	3.5			
	Persona reinsediata	Resettled person	1	0	0			
Totals			8		3.5	IVI = 0.4375	x0.00426	0.00186
RETURNEE	Rimpatriato	Returnee	22	0	0			
	Rimpatriando	Returnee	3	0	0			
Totals			25		0	IVI = 0		0
SEASONAL WORKER	Lavoratore stagionale	Seasonal worker	4	0	0			
Totals			4		0	IVI = 0		0
STATELESS PERSON	Apolide	Stateless person	33	0.5	16.5			
Totals			33		16.5	IVI = 0.50000	x0.01759	0.00886
THIRD-COUNTRY NATIONAL	Cittadino di paesi terzi	Third-country national	507	0	0			
	Cittadino di paesi terzi	Third country national	23	0	0			
	Cittadino di paesi terzi	National of third countries	31	0	0			
	Cittadino di paesi terzi	Non-EU national	4	0.5	2			
	Cittadino di paesi terzi	Thirdcountry national	5	0	0			
	Cittadino di paesi terzi	OMITTED	2	1	2			
	Cittadino di paesi terzi	National of non-EU countries	1	0.5	0.5			
	Cittadino di paesi terzi	TCN	1	0.5	0.5			
			574		5	IVI = 0.00871	x0.30597	0.00266

UNACCOMPANIED MINOR	Minore non accompagnato	Unaccompanied minor	45	0	0			
	Minore non accompagnato	UAM	1	0.5	0.5			
	Minore non accompagnato	Unaccompanied children	1	0.5	0.5			
	Minore rifugiato	Refugee child	7	0.5	3.5			
Totals			54		4.5	IVI = 0.08333	x0.02878	0.00240
UNION CITIZEN	Cittadino dell'Unione	Union citizen	13	0	0			
	Cittadino dell'Unione	Citizen of the Union	10	0	0			
	Cittadino dell'Unione europea	EU citizen	1	0.5	0.5			
	Cittadino dell'Unione europea	Citizen of the Union	1	0.5	0.5			
	Cittadino dell'Unione europea	Citizen of the European Union	1	0	0			
	Cittadino dell'UE	EU citizen	11	0	0			
	Cittadino dell'UE	EU national	3	0	0			
	Cittadino europeo	European citizen	5	0	0			
	Cittadino europeo	European	2	0.5	1			
	Cittadino europeo	EU citizen	1	0.5	0.5			
	Cittadino europeo	European people	1	0.5	0.5			
	Cittadino europeo	Citizen	1	0.5	0.5			
	Cittadino europeo	European national	1	0	0			
			51		3.5	IVI = 0.06863	x0.02719	0.00187
VICTIM OF TRAFFICKING IN HUMAN BEINGS	Vittima della tratta	Victims of trafficking	11	0	0			
	Vittima della tratta	Trafficked persons	2	0.5	1			
	Vittima della tratta	Trafficking	1	1	1			
	Vittima della tratta	Victims of human trafficking	1	0.5	0.5			
	Vittima della tratta di esseri umani	Victims of trafficking in human beings	24	0	0			
	Vittima della tratta di esseri umani	Victims of trafficking	7	0.5	3.5			
	Vittima della tratta di esseri umani	Victims of human trafficking	2	0	0			
	Vittima della tratta di esseri umani	Human trafficking victims	2	0	0			
			50		6	IVI = 0.12000	x0.02665	0.0032
VULNERABLE PERSON	Persona vulnerabile	Vulnerable person	15	0	0			
	Soggetto vulnerabile	Vulnerable person	1	0.5	0.5			

16	0.5	IVI = 0.03125	x0.008 53	0.00027	
				0.00255	Weighted mean IVI

Appendix 2: Full results tables for Stage 2: *MigrationInTheNews Corpus* analysis

Tables for conceptual categories ASYLUM SEEKER and REFUGEE

	Relative frequency (per million words)		% Change	Proportion of category	
	Earlier subcorpus	Later subcorpus		Earlier subcorpus	Later subcorpus
Asylum seeker	661.22	235.92	**-64.32%**[18]	58%	90%
Asylum-seeker	414.57	34.08	**-91.78%**	37%	10%
People seeking asylum	7.87	0	**N/A**	1%	0%
Applicant for asylum	7.87	0	**N/A**	1%	0%
Asylum applicant	41.98	0	**N/A**	4%	0%
Refugee	1,317.18	671.06	**-49.05%**	100%	100%

	Relative frequency (per million words)		% Change	Proportion of category	
	Earlier subcorpus	Later subcorpus		Earlier subcorpus	Later subcorpus
Demandeur d'asile	245.16	337.22	**37.55%**	100%	100%
Réfugié	606.55	989.33	**63.11%**	100%	100%

	Relative frequency (per million words)		% Change	Proportion of category	
	Earlier subcorpus	Later subcorpus		Earlier subcorpus	Later subcorpus
Richiedente protezione internazionale	0	15.46	**N/A**	0%	4%
Richiedente asilo	4.83	340.02	**6,939.75%**	100%	96%

[18] Values in bold in the % change column indicate those values that can be considered statistically significant according to the p-value derived from the log likelihood results

Rifugiato politico	21.73	20.61	-5.15%	5%	2%
Profugo	354.87	504.89	**42.27%**	76%	44%
Rifugiato	89.32	623.38	**597.92%**	19%	54%

Tables for conceptual category MIGRANT

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Immigrant	2,235.54	2,049.01	-8.34%	73%	52%
Migrant	325.36	1,860.11	**471.71%**	11%	47%
Entrant	443.43	15.74	**-96.45%**	15%	>1%
Emigrant	15.74	10.49	-33.35%	1%	>1%
Newcomer	31.49	31.48	-0.03%	1%	1%

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Migrant	141.54	1,450.52	**924.81%**	9%	53%
Expatrié	0	7.44	**N/A**	>1%	>1%
Immigré	692.51	939.74	**35.70%**	46%	34%
Immigrant	588.89	260.35	**-55.79%**	39%	10%
Nouvel arrivant	48.02	69.43	44.59%	3%	3%
Emigrant	22.75	0	**N/A**	2%	0%

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Migrante	21.73	1,620.27	**7,356.37%**	1%	56%
Migrato	0	7.73	**N/A**	0%	>1%
Immigrato	2,013.34	1,241.61	**-38.33%**	91%	43%
Immigrante	53.11	15.46	**-70.89%**	2%	1%
Immigrato extracomunitario	31.38	0	**N/A**	1%	0%
Emigrante	55.52	0	**N/A**	3%	0%
Emigrato	28.97	15.46	-46.63%	1%	1%

Tables for conceptual category IRREGULAR MIGRANT

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**

Illegal immigrant	750.43	545.70	**-27.28%**	61%	73%
Illegal entrant	354.22	0	**N/A**	29%	0%
Illegal	65.60	31.48	**-52.01%**	5%	4%
Illegal alien	44.61	0	**N/A**	2%	0%
Illegal migrant	34.11	104.94	**207.65%**	3%	14%
Undocumented immigrant	0	10.49	**N/A**	0%	1%
Irregular immigrant	0	18.36	**N/A**	0%	2%
Undocumented migrant	0	20.99	**N/A**	0%	3%
Unauthorised migrant	0	7.87	**N/A**	0%	1%
Irregular migrant	0	5.25	N/A	0%	>1%

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Etranger sans papiers	10.11	0	**N/A**	1%	0%
Etranger clandestin	7.58	0	**N/A**	1%	0%
Etranger en situation irreguliere	103.62	109.10	5.29%	8%	16%
Etranger en situation illegale	10.11	12.40	22.65%	1%	2%
Personne en situation irreguliere	15.16	27.27	79.88%	1%	4%
Clandestin	510.54	250.43	**-50.95%**	39%	34%
Sans-papiers	444.83	104.14	**-76.59%**	34%	14%
Immigre en situation irreguliere	10.11	0	**N/A**	1%	0%
Immigre clandestin	48.02	0	**N/A**	4%	0%
Immigre illegal	17.69	19.84	12.15%	1%	3%
Immigrant illégal	45.49	14.88	**-67.29%**	3%	2%
Immigrant clandestin	30.33	0	**N/A**	2%	0%
Un illegal	48.02	61.99	29.09%	4%	9%
Un irregulier	17.69	0	**N/A**	1%	0%

Migrant illégal	0	24.80	**N/A**	0%	3%
Migrant irrégulier	0	9.92	**N/A**	0%	1%
Migrant sans papiers	0	14.88	**N/A**	0%	2%
Migrant clandestin	0	19.84	**N/A**	0%	3%
Migrant en situation irrégulière	0	17.36	**N/A**	0%	2%
Étranger entré illégalement	0	19.84	**N/A**	0%	3%
Immigrant sans papiers	0	9.92	**N/A**	0%	1%
Immigrant en situation irrégulière	0	7.44	**N/A**	0%	1%

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Clandestino	939.07	208.65	**-77.78%**	81%	47%
Immigrato irregolare	19.31	20.61	6.73%	2%	5%
Immigrato clandestino	113.46	28.34	**-75.02%**	10%	6%
Immigrato illegale	16.90	25.76	**52.43%**	1%	6%
Clandestino extracomunitario	12.07	0	**N/A**	1%	0%
Extracomunitario clandestino	7.24	0	**N/A**	1%	0%
Extracomunitario irregolare	9.66	0	**N/A**	1%	0%
Extracomunitario non in regola	9.66	0	**N/A**	1%	0%
Sans-papiers	26.55	0	**N/A**	2%	0%
Migrante irregolare	0	12.88	**N/A**	0%	3%
Straniero irregolare	0	15.46	**N/A**	0%	3%
Irregolare	0	95.31	**N/A**	0%	22%
Illegale	0	28.34	**N/A**	0%	6%
Immigrante illegale	0	7.73	**N/A**	0%	2%

Tables for conceptual category THIRD-COUNTRY NATIONAL

	Relative frequency (per million words)		% Change	Proportion of category	
	Earlier subcorpus	Later subcorpus		Earlier subcorpus	Later subcorpus
Foreigner	362.09	128.55	**-64.50%**	57%	50%
Alien	204.66	10.49	**-94.87%**	32%	4%
Outsider	23.61	0	**N/A**	4%	0%
Person from abroad	15.74	0	**N/A**	2%	0%
Foreign-born person	0	10.49	**N/A**	0%	4%
Person from overseas	0	7.87	**N/A**	0%	3%
Foreign citizen	0	13.11	**N/A**	0%	5%
Non-Europeans	0	15.74	**N/A**	0%	6%
Non-EU citizen	0	28.86	**N/A**	0%	11%
Foreign national	26.24	41.98	**59.98%**	4%	16%

	Relative frequency (per million words)		% Change	Proportion of category	
	Earlier subcorpus	Later subcorpus		Earlier subcorpus	Later subcorpus
Etranger	1,670.62	1,202.57	**-28.02%**	98%	94%
Personne d'origine etrangere	10.11	7.44	-26.41%	1%	1%
Ressortissant etranger	15.16	0	**N/A**	1%	0%
Ressortissant de pays tiers	5.05	14.88	194.65%	>1%	1%
Étranger en provenance de pays tiers	0	12.50	**N/A**	0%	1%
Étranger non européen	0	12.50	**N/A**	0%	1%
Personne née à l'etranger	0	14.88	**N/A**	0%	1%
Personne étrangère	0	9.92	**N/A**	0%	1%

	Relative frequency (per million words)		% Change	Proportion of category	
	Earlier subcorpus	Later subcorpus		Earlier subcorpus	Later subcorpus

Straniero	726.64	659.44	-9.25%	50%	88%
Cittadino straniero	28.97	41.22	42.29%	2%	6%
Extracomunitario	663.87	30.91	**-95.34%**	45%	4%
Cittadino extracomunitario	36.21	15.46	-57.30%	2%	2%
Non europeo	7.24	0	**N/A**	>1%	0%

Tables for conceptual category UNION CITIZEN

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
EU citizen	23.61	396.92	**1,581.15%**	26%	68%
European	49.85	149.54	**199.89%**	54%	28%
European Union citizen	7.87	0	**N/A**	9%	0%
European citizen	5.25	20.99	**299.81%**	6%	4%
National of member states	5.25	0	**N/A**	6%	0%

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Ressortissant de l'union europeenne	7.58	0	**N/A**	15%	0%
Ressortissant européen	0	14.88	**N/A**	0%	5%
Citoyen européen	2.53	32.23	**1,173.91%**	5%	11%
Européen	37.91	247.95	**554.05%**	75%	83%
Ressortissant communautaire	2.53	0	N/A	5%	0%
Citoyen de l'UE	0	4.96	N/A	0%	2%

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Cittadino CEE	9.66	0	**N/A**	17%	0%
Cittadino comunitario	0	15.46	**N/A**	0%	10%
Cittadino dell'UE	0	7.73	**N/A**	0%	5%

Cittadino europeo	0	23.18	**N/A**	0%	15%
Cittadino UE	0	15.46	**N/A**	0%	10%
Europeo	48.28	92.73	**92.07%**	83%	59%
Cittadino dell'Unione Europea	0	2.58	N/A	0%	2%

Tables for conceptual category MIGRANT WORKER

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Economic migrant	68.22	18.36	**-73.09%**	59%	6%
Immigrant worker	20.99	7.86	-62.55%	18%	3%
Migrant worker	15.74	68.21	**333.36%**	14%	23%
Foreign worker	10.50	102.32	**874.48%**	9%	35%
Immigrant labour	0	13.12	**N/A**	0%	4%
Overseas worker	0	23.61	**N/A**	0%	8%
Migrant labour	0	26.24	**N/A**	0%	9%
Non-EU worker	0	34.11	**N/A**	0%	12%

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Travailleur étranger	40.44	14.88	**-63.20%**	33%	13%
Travailleur immigré	50.55	27.27	-50.94%	39%	24%
Travailleur clandestin	25.27	4.96	**-80.37%**	20%	4%
Employé étranger	0	9.92	**N/A**	0%	9%
Salarié étranger	0	14.88	**N/A**	0%	13%
Travailleur migrant	5.05	4.96	-1.78%	4%	4%
Migrant économique	5.05	27.27	**440.00%**	4%	24%

Immigrant économique	0	7.44	**N/A**	0%	7%

	Relative frequency (per million words)		**% Change**	**Proportion of category**	
	Earlier subcorpus	**Later subcorpus**		**Earlier subcorpus**	**Later subcorpus**
Lavoratore straniero	36.21	25.76	-28.86%	60%	26%
Lavoratore extracomunitario	14.48	0	**N/A**	24%	0%
Lavoratore immigrato	9.66	0	**N/A**	16%	0%
Migrante economico	0	64.4	**N/A**	0%	66%
Immigrante economico	0	7.73	**N/A**	0%	8%

Appendix 3: Stage 3: Proportion of concepts per subcorpora

	Eng EUML [19] 1992-1998	Eng MITN [20] 1992-1998	Eng EUML 2015-2018	Eng MITN 2015-2018	Fr EUML 1992-1998	Fr MITN 1992-1998	Fr EUML 2015-2018	Fr MITN 2015-2018	It EUML 1992-1998	It MITN 1992-1998	It EUML 2015-2018	It MITN 2015-2018
ASYLUM SEEKER	**11.11 %**	**14.56 %**	**5.27%**	**3.02%**	**10.84 %**	**4.32%**	**5.28%**	**5.07%**	**10.79 %**	0.09%	**6.34%**	**6.01%**
CROSS-BORDER WORKER	0.37%	0%	0.44%	0%	0%	0%	0.46%	0%	0%	0%	0.35%	0%
DETAINEE	0%	0.61%	0%	0.26%	0%	0.13%	0%	0%	0%	0%	0%	0%
DISPLACED PERSON	**2.06%**	0%	0.83%	0%	**1.79%**	0.09%	**1.38%**	0.15%	**1.99%**	**2.16%**	**1.41%**	0.31%
INTERNATIONAL STUDENT	0%	0.51%	0%	**2.98%**	0%	0.40%	0%	0.41%	0%	0%	0%	0%
IRREGULAR MIGRANT	**4.47%**	**16.04 %**	**3.22%**	**10.67 %**	**4.80%**	**23.27 %**	**3.02%**	**10.89 %**	**4.84%**	**20.68 %**	**3.05%**	**7.49%**
LONG-TERM RESIDENT	**1.83%**	0%	0.17%	0%	**2.12%**	0%	0.20%	0%	**1.86%**	0%	0.23%	0%
MIGRANT	**25.54 %**	**39.18 %**	**24.53 %**	**56.40 %**	**21.56 %**	**26.35 %**	**24.69 %**	**41.01 %**	**21.59 %**	**39.51 %**	**25.66 %**	**49.06 %**
MIGRANT WOMAN	0.46%	0%	0%	0%	0.56%	0%	0%	0.11%	0.87%	0.13%	0%	0.13%
MIGRANT WORKER	**4.15%**	**1.48%**	0.83%	**4.18%**	**3.13%**	**2.18%**	0.46%	**1.60%**	**4.34%**	**1.08%**	0.81%	**1.66%**
MOBILE WORKER	0%	0%	0.78%	0%	0%	0%	0.67%	0%	0%	0%	0.82%	0%
OFFSHORE WORKER	0%	0%	0.22%	0%	0%	0%	0.20%	0%	0%	0%	0.18%	0%
OVERSTAYER	0%	0.84%	0%	0.34%	0%	0%	0%	0%	0%	0%	0%	0%
POSTED WORKER	0%	0%	0.55%	0%	0%	0%	0.26%	0.19%	0%	0%	0.53%	0%
REFUGEE	**11.68 %**	**16.91 %**	**21.20 %**	**8.58%**	**12.96 %**	**11.28 %**	**20.34 %**	**14.65 %**	**15.88 %**	**8.35%**	**23.43 %**	**19.43 %**
REJECTED APPLICANT FOR INTERNATIONAL PROTECTION	0%	0.17%	0%	0.48%	0.22%	0.53%	0%	**1.75%**	0%	0%	0%	0%
RELOCATED PERSON	0%	0%	0.39%	0%	0%	0%	0.26%	0%	0%	0%	0.23%	0%
RESETTLED PERSON	0%	0%	0.11%	0%	0%	0%	0.51%	0%	0%	0%	0.47%	0%
RETURNEE	0.46%	0.47%	**1.55%**	0%	0%	0%	0.82%	0%	0%	0.22%	**1.94%**	0%
SEASONAL WORKER	0.57%	0%	0.11%	0.15%	0.67%	0.13%	0.20%	0.11%	0.74%	0.36%	0.23%	0.04%
STATELESS PERSON	0%	0%	**1.89%**	0%	0%	0%	**1.79%**	0.26%	0%	0%	**2.06%**	0%
THIRD-COUNTRY NATIONAL	**26.46 %**	**8.12%**	**30.58 %**	**3.66%**	**30.39 %**	**30.00 %**	**29.41 %**	**19.16 %**	**25.68 %**	**26.22 %**	**22.67 %**	**12.64 %**
UNACCOMPANIED MINOR	0.46%	0%	**2.83%**	0.41%	0.67%	0.04%	**2.66%**	0.19%	0.99%	0%	**2.94%**	0.26%
UNION CITIZEN	**7.79%**	**1.18%**	**2.72%**	**7.68%**	**8.27%**	0.89%	**2.20%**	**4.81%**	**9.55%**	**1.25%**	**1.76%**	**2.66%**
VICTIM OF TRAFFICKING IN HUMAN BEINGS	**3.21%**	0%	**2.55%**	0%	**3.80%**	0%	**3.23%**	0%	**2.73%**	0.04%	**3.58%**	0%
VULNERABLE PERSON	0%	0%	0.94%	0.48%	0%	0%	0.82%	0.04%	0%	0%	0.88%	0%

[19] EUML = *EUMigrLaw Corpus*

[20] MITN = *MigrationInTheNews Corpus*