

Masterarbeit zur Erlangung des akademischen Grades **Master of Arts** der Philosophischen Fakultät der Universität Zürich

# Corpus Linguistics Approaches for Multilingual Discourse Analysis

A semi-automatic comparative analysis of the Swiss-German and the Italian political discourse about immigration

Verfasserin: Michela Rossi

Matrikel-Nr: 10-736-072

Referent: Prof. Dr. Martin Volk

Betreuer: Dr. Noah Bubenhofer

Institut für Computerlinguistik

Abgabedatum: 10.07.2016

## Abstract

In this project we combined corpus linguistics (CL), statistical machine translation (SMT) and discourse analysis (DA). The goal of this interdisciplinary thesis consists in showing how this combination can be fruitful for DA, by performing a comparison in the political discourse about immigration between the German-speaking part of Switzerland and Italy.

Using CL approaches to statistically retrieve linguistic elements, significant for a specific topic, proved to be productive in the last years. These methods allow to identify concrete data by calculating discourses. These can afterwards be classified and re-conducted to the traditional analysis categories of DA, such as topoi, metaphors, metonymies, etc. In this manner it is the corpus itself that provides solid bases to generate hypothesis about the analysed portion of text. The challenge of this investigation consisted in retrieving these linguistic elements, and then use them to directly compare the two discourses about immigration through SMT. This was done by training an SMT system with a big German-Italian parallel corpus.

Consequently, we could discover similarities and particularly differences between the two discourses, outlining the argumentative strategies mainly implemented. A multilingual analysis is proven to be very useful to investigate a controversial and contemporary topic, which overcomes national and linguistic borders, such as immigration.

## Zusammenfassung

In diesem Projekt haben wir Methoden der Korpuslinguistik (KL) mit statistischer Maschineller Übersetzung (SMT) und Diskursanalyse (DA) in Verbingung gesetzt. Ziel dieser interdisziplinären Arbeit ist es zu zeigen, inwiefern diese Kombination für DA fruchtbar sein kann, indem ein Vergleich im politischen Diskurs über das Thema 'Migration' zwischen der Deutsch-Schweiz und Italien durchgeführt wird.

KL Vorgehensweisen, die statistisch themenspezifische linguistische Elemente extrahieren, haben sich in den letzten Jahren als sehr produktiv etabliert. Diese Methoden ermöglichen Diskursberechnungen, indem konkrete Elemente identifiziert werden, die dann in einem zweiten Schritt zu den traditionellen Analysekategorien der DA, wie beispielsweise Topoi, Metaphern oder Metonymien, zurückgeführt werden können. Dementsprechend, handelt es sich um das Korpus selbst, das solide Grundlagen für die Hypothesenbildung zur Verfügung stellt. Das spannende an dieser Erforschung besteht in der Extraktion dieser Elemente, die dann zu einem direkten Vergleich zwischen den zwei politischen Diskursen benuzt werden. Dieser Vergleich wird durch das Trainieren eines SMT System mit einem grossen Deutsch-Italienischen Parallel Korpus erzeugt.

Damit konnten wir rausfinden welche Ahnlichkeiten und vor allem Unterschiede zwischen den zwei Diskursen bestehen, und die am meisten benutzten Argumentationsstrategien identifizieren. Einen multilingualen Ansatz zeigt sich als sehr hilfsreich in der Analyse eines gegenwärtigen und umstrittenen Themas, das nationale und linguistische Grenzen überschreitet, wie das der Migration.

# Acknowledgement

I want to thank all the people who supported me with this Master's thesis.

Firstly, I thank Dr. Noah Bubenhofer for the good supervision of the project and for his guidance in the field of Corpus Linguistics through many explanations and advices. Furthermore, I would like to thank Prof. Martin Volk for providing me the *bulletin4corpus* and for supporting the project. Many thanks to Stefano Pongelli and to my father Armando Rossi for proofreading the present text. I should also thank Sara Groisman, Felicity Brunner, and Sara Wick for giving me some advices regarding single sections of the paper. I also want to express special thanks to Chiara Baffelli, with which I shared all my educational path in the filed of MLTA. Moreover, heartfelt thanks to Davide Arrigo, who always supported me in both this project and all my studies. Finally, I thank my friends and family for all their love and affection.

# Contents

Ab	ostrac	ct																		i
Ac	knov	vledgeme	nt																	iii
Co	onten	ts																		iv
Lis	st of I	Figures																		vi
Lis	st of <sup>-</sup>	Tables																		vii
Lis	st of <i>i</i>	Acronyms	;																	ix
1	Intro	oduction																		1
	1.1	Motivati	on							•		 •		•						1
	1.2	Research	Question	s											•					2
	1.3	Thesis St	ructure.						•	•	•	 •	 •	•	•	•	•	•	•	4
2	Disc	ourse An	alysis																	<b>5</b>
	2.1	Political	Discourse							•					•					7
	2.2	Analysis	Categorie	s					•	•••		 •	 •		•	•				7
	2	2.2.1 Leg	itimation						•	•		 •			•					8
	2	2.2.2 Co	ercion						•	• •		 •	 •		•					9
		2.2.2.1	Ingroup	vs. Or	utgro	up.			•	•••	•		 •	•		•				11
		2.2.2.2	Emotive	e Coerc	ion .				•	•••	•			•	•	•			•	13
		2.2.2.3	Cognitiv	ve Coei	rcion				•		•	 •	 •	•	•	•	•	•	•	15
3	Cor	pus Lingu	istics																	18
	3.1	Corpus-I	Priven and	l Corpi	us-Ba	sed									•					19
	3.2	Extractio	on of Sign	ificant	Elem	ents									•					20
	3	8.2.1 Ke	ywords Ex	ctractic	m				•	•	•	 •		•	•					21
	3	3.2.2 Co-	occurrenc	es and	MW	Es E	Extr	act	ion		•	 •		•		•	•	•	•	23

4	Statistical	Machine	Translation
---	-------------	---------	-------------

26

5	Corpora	<b>28</b>
	5.1 Primary Political Corpora	28
	5.1.1 Corpora Construction	29
	5.1.2 Immigration Sub-corpora	30
	5.2 Translation Corpora	32
	5.2.1 Europarl $\ldots$	32
	5.2.2 bulletin4corpus $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	33
6	Calculating Discourses	35
	6.1 Preprocessing	35
	6.2 Extracting Keywords	37
	6.3 Extracting Co-ocurrences	39
	6.4 Extracting Multi-Words-Expressions	41
7	Comparing Discourses	43
	7.1 Preprocessing	44
	7.2 Translation Model	44
	7.3 Keywords Comparison	45
8	Analysis and Interpretation	50
	8.1 Thematic Comparison	51
	8.2 Legitimation	53
	8.3 Coercion	56
	8.3.1 Ingroup Vs. Outgroup	56
	8.3.2 Emotive Coercion	61
	8.3.3 Cognitive Coercion	68
	8.4 Observations	77
9	Conclusion	80
	9.1 Limits and Future Prospects	81
Re	eferences	83
Le	ebenslauf	87
Δ	Selection of Significant Extracted Elements	88
	A.1 Keywords	89
	A.2 Intersecting Keywords	91
	A.3 3-Grams, 4-Grams, 5-Grams	92
в	List of Corpora and Tools	97

# **List of Figures**

1	X Critical Value	22
2	bulletin4corpus	33
3	bulletin4corpus-Alignmentfile	34

# **List of Tables**

1	General Contingency Table	21
2	Co-occurrences Contingency Table	24
3	Corpus Information	29
4	Swiss-German Parties	29
5	Italian Parties	30
6	German and Italian most Significant Keywords	38
7	Grouped Noun Co-occurrences	40
8	Significant 4-Grams	42
9	Lexfile Output	45
10	Translation File Italian to German	46
11	Translation Match Italian to German	47
12	Intersecting Keywords	48
13	General Terms	51
14	EU Migrant Crisis	52
15	EU Intern Migration	53
16	Co-occurrences of <i>Politiker</i>	54
17	Possessive Nouns	56
18	Referential Metonymies	58
19	Swiss Positive Face Addressing	60
20	Italian Negative Face Mistreating	60
21	Metaphors	62
22	Italian Topos of Danger	64
23	Italian Topos of Diseases	65
24	Italian Topos of Death	66
25	Swiss-German Topos of Benefits	69
26	Swiss-German Topos of Disadvantage	69
27	Italian Topos of Displacement	71
28	Italian Topos of Inverse Racism	72
29	Swiss-German Topos of Exploitation	73
30	Swiss-German Topos of Crime	75
31	Swiss-German Topos of Culture	76
32	Swiss-German Topos of Overload	77

33	German Keywords	89
34	Italian Keywords	90
35	Intersecting Keywords	91
36	German 3-Grams	92
37	Italian 3-Grams	93
38	German 4-Grams	94
39	Italian 4-Grams	95
40	German 5-Grams	96
41	Italian 5-Grams	96

# List of Acronyms

$\operatorname{CL}$	Corpus Linguistic
DA	Discourse Analysis
EM	Expectation Maximum Algorithm
MT	Machine Translation
LM	Language Model
MWE	Multi-Word-Expression
POS	Part-Of-Speech
SMT	Statistical Machine Translation
ТМ	Translation Model
XML	eXtensible Markup Language

# 1 Introduction

## 1.1 Motivation

Traditional discourse analysis (DA) has always relied on qualitative examination of text corpora as its empirical foundation. Nowadays, Corpus Linguistics (CL) approaches, in particular corpus driven ones, employ statistical and contrastive methods in order to improve DA. The corpus-driven perspective aims to outline linguistic patterns, which are typical for a determinate text portion, by working inductively (Bubenhofer [2008], Bubenhofer [2009]).

Foucault described discourses as practices representing and determining interpersonal configurations, as well as object of knowledge of the society (Foucault [1972]). Discourses possess the double attribute of being influenced by society and to influence it at the same time. Political discourses, whose objective is to influence text consumers, typically involve persuasion and bargaining (Chilton [2004]:4). These actions are implemented concretely by argumentative strategies built upon metaphors, topoi, metonymies, and reasoning patterns: figures that contribute to affect text consumers at both emotive and cognitive level. A corpus driven approach seeks exactly to identify these linguistic structures. Statistically retrieving significant elements (keywords, co-occurrences, and multi-word-expressions) typical of a determinate topic by relying exclusively on real data eliminates, furthermore, the risk of being influenced by preconceptions. We applied this approach to a contemporary and discussed topic, investigating immigration in political discourse.

Moreover, statistical machine translation (SMT) offers the possibility to compare two discourses belonging to different countries in different languages by directly comparing the translations of the retrieved keywords of both discourses. Therefore, we decided to investigate similarities and differences in the argumentative strategies of the immigration discourse, in the German-speaking part of Switzerland and Italy. We did this expecting to find mainly differing elements, taking into consideration geographical, historical, and socio-economical differences existing between the two countries. Switzerland is historically a multilingual nation, in which four distinct cultures and identities coexist. In this investigation we decided to focus on the German speaking part of the nation, which reflects the largest ethnic group of the country, in order to limit the multilingual comparison to the language pair German-Italian.

Many research studies were conducted by combining CL methods with DA. Most of them following a corpus-based approach, namely starting from a set of specific terms and concepts and using different CL methods to investigate their frequency, co-occurrences, and syntactic patterning in diverses specific corpora (Orpin [2006],Mayaffre and Poudat [2013]). One example of study that combines the corpus-driven and corpus-based approach was performed by Bubenhofer, in a diachronic investigation about the foreign affairs section of the NZZ<sup>1</sup> (Bubenhofer [2009]:209-293).

In our investigation we adopted a synchronic point of view, combining a corpusdriven/corpus-based approach with SMT. This gave us the possibility to perform a multilingual and multicultural semi-automatic discourse comparison between two specific discourses broadening the research prospective in that field. As already mentioned, we choose Switzerland and Italy because of the profound differences at historical, socio-economical, and cultural level. Furthermore, the two countries differ regarding the specific topic of immigration. Switzerland is among the European countries with the largest immigrants population; 27% of the full aged people are foreign-born and the immigration issue is often discussed in the political agenda, as well as in the public debate (Thomas et al. [2012]:5). Italy, on the contrary, is traditional an emigration country that has recently become an immigration one (Triandafyllidou [2000]:373). These socio-historical facts provide the foundation for interesting differences at argumentative level in the political discourse.

## **1.2 Research Questions**

In this project we firstly created two political corpora with the open source tool import.io<sup>2</sup>. The corpora contained press releases of the main political parties of German-speaking Switzerland and Italy between 2006 and 2016. Afterwards, we isolated articles concerning the topic of immigration from both corpora. This was done in order to contrastively extract significant elements in a further step. The contemporary large digital data availability allows to create topic-specific corpora

<sup>&</sup>lt;sup>1</sup>Neue Zürcher Zeitung.

<sup>&</sup>lt;sup>2</sup>https://www.import.io/.

large enough to be representative (see chapter 5).

In an additional step (see chapter 6) we performed the discourse calculation, by contrastively extracting keywords, co-occurrences, and MWEs. The extractions were implemented with contingency tables and statistical calculations such as Chisquared test and Log-likelihood. We choose Python<sup>3</sup> as programming language to execute the computations. Keywords were extracted in order to analyse the main concepts that emerged as typical for the immigration topic. Co-occurrences were useful to investigate the conceptual association of keywords in their context. Finally, MWEs, which are the best indicators of linguistic crystallisation processes, have proven to be determinant for the recognition of cognitive patterns.

The next step (see chapter 7) consisted in a direct discourse comparison through SMT. We trained the Translation Model (TM) of an SMT system with Moses<sup>4</sup>. The training corpus consisted of German-Italian aligned articles from *Europarl*<sup>5</sup> and *bulletin4corpus*<sup>6</sup>. Both corpora were used to obtain a combined corpus about the immigration topic in the German-speaking part of Switzerland and Italy.

Extracted elements and translation system provided the foundations to perform a comparative analysis of the argumentative strategies in use in the political discourse about immigration of the two countries (see chapter 8).

The research questions of this thesis are articulated on two different operational levels: methodology and contents. The main research questions, which shall be answered regarding the implemented methodology, are:

- In what way does traditional DA benefit from statistical CL methods?
- Is it possible to compare two discourses of different languages and cultures by combining this CL approach with SMT?

Concerning the contents level we will answer the following question:

- Which cognitive structures, metaphors, and topoi characterise the Swiss-German and the Italian political discourse about immigration?
- Do the discourses show similarities and/or differences?

<sup>&</sup>lt;sup>3</sup>https://www.python.org/.

<sup>&</sup>lt;sup>4</sup>http://www.statmt.org/moses.

<sup>&</sup>lt;sup>5</sup>http://www.statmt.org/europarl/.

<sup>&</sup>lt;sup>6</sup>http://kitt.cl.uzh.ch/kitt/b4c/en/.

## **1.3 Thesis Structure**

In this paper we present our interdisciplinary project which, combining the fields of CL, SMT, and DA, aims to investigate the political discourse about immigration in the German-speaking part of Switzerland and Italy by calculating and comparing it. The thesis consists of nine chapters.

In the introduction we provide a general overview of the project, introducing motivation, objectives and precise research questions. The following three chapters introduce the different fields constituting the theoretical background of this interdisciplinary research. In chapter 2 we present fundamental theories of DA, focusing mainly on the analysis categories that we are going to use in the chapter about interpretation and analysis of the results (see chapter 8). Chapter 3 introduces the theoretical field of CL, the heart of this section consists in the exemplification of statistical methods to compute and extract keywords, co-occurrences, and MWEs. Chapter 4 provides an introduction to SMT, focusing on the TM. In chapter 5 we describe and discuss the composition of both corpora types: primary and translation ones. We used primary corpora to extract significant elements, and translation corpora to perform the discourses comparison. Chapter 6 consists in the description of the actual discourse computation procedure which was divided in a pre-processing step, followed by the extraction of keywords, co-occurrences, and the MWEs. As next steps we describe the SMT training and then the procedure with which we compared the two previously extracted keywords-lists (see chapter 7). In chapter 8 we present the analysis and interpretation of the results. This is primarily based on the keywords comparison, integrated with both co-occurrences of these keywords in their context and MWEs when available. Finally, in chapter 9, we summarise the whole project in light of the obtained results, answer research questions, and discuss limits of the approach and future prospects.

# 2 Discourse Analysis

In this chapter we will introduce the first big field of the theoretical background of our investigation: discourse analysis (DA).

The first step to approach DA is to clear the term *discourse* and to understand it in its historical perspective. The term *discourse* is unclear and above all associated with an enormous amount of senses and concepts. This ambiguity is due, as Warnke (2007:3) noticed, to its various usages to express different meanings in everyday and scientific language. Furthermore, *discourse*, was associated with different concepts in the European language history, being related to a number of research directions like literature, language, and communication studies.

In linguistics, *discourse* appears for the first time in relation to the American structuralism (Warnke [2007]:4). The linguist Zellig Harris (1952) is the first to express the necessity to analyse the connection between linguistic elements overcoming the sentences borders, in the so-called connected speech. Harris understands the *discourse* term as the modern *text* term: a sentence crossing linguistic act. The positive reception to Harris' theory leads to the establishment of more text-related definitions of *discourse* in the linguistic field. The oral expression is not excluded from this definition and becomes more and more important especially in the 80's and early 90's. In this context, and based on this single text referring definition of *discourse* intended as the language level of a specific production related to a specific speaker in a specific context (Brown and Yule [1983]:9), different linguists, like Brown and Yule, theorise their approach to DA.

In other, non-linguistic, disciplines the term *discourse* is more or less always used in reference to Michel Foucault. Foucault's *discourse* concept strongly contrasts the idea of *discourse* as a single text referring entity. To summarise this concept in a massively simplified way: *discourse* is understood as a collection of linguistic expressions which are semantically related, and used by a particular group of society (Foucault [1972]). *Discourse* is not considered as a passive mirror of a given reality, but more as a practice which determines precise interpersonal configurations and objects of knowledge. Communicative events are considered in relation with the communicative situation, with other discourses that interact with them, and with the location and time of their production. This approach became fruitful in various disciplines such as sociology, literature, and philosophy. Consequently, the usage of Foucault's *discourse* in other disciplines brought, through a process that involved the reception of Roland Barthes, Julia Kristeva, Michel Pecheux, Gerarde Genette, and Tzevetan Todorov, to an enlargement of the linguistic terminology spectrum (Warnke [2007]:4).

The contemporary *discourse* term is based on Foucault's discourses not only in sociology, literature, and philosophy, but even in the linguistic field. To clarify this modern term definition, we can use the following four elements identified by Gardt (2007:29-30):

- 1. Discourse as concatenation of different texts;
- 2. Pragmatical orientation of a discourse;
- 3. Connection between discourse and society, i.e. discourse seen as expression of a society or of a group that is part of it;
- 4. Discourse as input for social change, perceived as entity that not only represent the society, but also actively contribute to its mental and ontological establishment;

Summarising, in this investigation we will consider *discourse* as a reflection about one argument that can include every kind of expressions and texts, fulfilled by a determinate group of people. *Discourse*, furthermore, reflects the attitude and knowledge of this group regarding the argument, but also actively contribute to its grounding in the future (Gardt [2007]:30).

In light of this clarification we can now move on to the concept of DA. Antelmi (2012:X) defines it as a strategy, oriented to Foucault's research, which moves away from the anxiety generated by the proliferation of discourses in the history of occidental society. In order to perform a pragmatical/functional understanding of discourses, this strategy, that focuses on every level of the language system, is mostly concentrated on the two levels of lexicon and system. The linguistic surface of different texts presents, in fact, a discursive connection (Bubenhofer [2008]:408). DA aims ultimately to understand this discursive connection, by investigating its basic semantic shapes and deep semantic structures, which reflect determinate mental and thought patterns. These have a notable importance since, as said, on one hand they reflect the society and their power relations, and on the other hand, they contribute to configure exactly these relations.

## 2.1 Political Discourse

Interactions and relationships have always characterized every human community. For this reason, already in ancient times Aristotle described humans as "political animals" that live in *polis* (Greek word for state), regulated by these relationships. Consequently, politics studies often explained politics in relations to the entity of power (Schäffner and Bassnett [2010]:2). As Chilton (2004:3) points out, the struggle to power between persons who aim to maintain their power and those who seek to resist it is a central issue in politics. At the same time politics is viewed as cooperation, which is needed to solve clashes of interest over influence, money, and interaction.

Language determines the main part of human interaction, which is influenced by socio-cultural, ideological, historical, and institutional conditions. It follows, therefore, that politics is deeply bound with language as well as with the political discourse, which consist of a complex form of human interaction in discursive practice (Schäffner and Bassnett [2010]:3).

Following the definition of van Dijk (1997:2), political discourse is primarily intended on the basis of its actors or authors: the politicians. Indeed, it consists of the amount of text and talks produced by political institutions or professional politicians. With "politicians" van Dijk intends the group of people who are on one hand paid for their political activity, and play, on the other hand, an important role as central placers in the political scene. This definition based on the actors is not very exhaustive as van Dijk points out. We should take into account many more other variables. In fact, "as soon as we locate politics and its discourses in the public sphere, many more participants in political communication appear on the stage" (Teun A. [1997]:13). The definition should also include the other participants: audience, voters, dissidents, and demonstrators, as well as the context of communication.

For our research we will limit ourselves, as we will explain in section 5, to considering the political discourse as the amount of news and press releases published by the most representative political parties on their websites.

## 2.2 Analysis Categories

In this section we will describe and discuss the main analysis categories that we are going to use in order to perform our final analysis. We will structure our analysis staring from two theoretical work: Analysing Political Discourse. Theory and Practice from Chilton (2004) and Critical Discourse Analysis and Cognitive Science. New Perspective on Immigration Discourse from Hart (2010). We will then integrate their observations with further analysis categories.

Coercion has been identified as the central strategy of the argumentative discourse. Chilton defines it as "an intention to affect the beliefs, emotions, and behaviours of others in such a way that suits one's interest" (2004:88). This is usually performed in argumentative discourse about immigration with specific linguistic strategies that oppose a positive represented ingroup (to which the text is addressed) to a negative constructed outgroup.

The first necessary strategy to perform coercion is legitimation. As Hart noticed, successful coercion does strictly depend on it (2010:89). Legitimation consists in establishing the right to be listened and obeyed by the text consumers (Chilton [2004]:111). From a cognitive point of view, the communication recipient is free to react in two different ways to a representation arriving from an argumentative text. He can accept it as true and store it in long time memory, influencing his own mental representation, or he can refuse it and forget it. Trough successful legitimation, the text consumer is more likely to accept the representation. Hart furthermore identifies, above generic coercion, two different specific kinds of coercion typical of the discourse about immigration: emotive and cognitive coercion. In the first kind of coercion the text producers try to use linguistic processes that evoke determinate emotions to perform coercion. The second kind, on the other hand, involves producing cognitive effects in text consumers by influencing their mental representation about a subject. To put these strategies into practice, mental processes such as metaphors and metonymies (Saiz de Lobado and Bonomi [2012]), topoi/enthymeme (Wengeler [2003b], Hart [2010]) and face theory (Goffman [1989]) which have already been theorised and studied by a number of researchers, are used. We will now analyse each strategy with its mental processes more closely.

#### 2.2.1 Legitimation

As previously mentioned, legitimation is vital in any argumentative strategy. The right to be listened is communicated through the use of a certain language over both explicit statements and implications. This strategy mostly includes arguments about willing of voters, general ideological principles, and charismatic leadership projection made through positive self-representation. Chilton identifies one main strategy to perform legitimation in political texts which he defines as epistemic strategy. This

consists in the claim of the text producer to have better knowledge as well as a better recognition of the "real" fact than the text consumers. The text producer presents himself as more objective and rational than normal citizens, establishing a sort of super instance. Statistics, numbers, and sources that are to be accepted as authoritative often support epistemic claims.

#### 2.2.2 Coercion

Coercion has its roots in the self and other representation. This dichotomous relation expects a positive connoted ingroup versus a negative connoted outgroup. It is achieved through the use of two sub-strategies: *reference* and *predication*. The *reference* sub-strategy consists in the linguistic construction used to concretely reference members of the ingroup and more often of the outgroup, achieving cognitive effects by strengthening the dichotomous representation we - they. Furthermore, it is often used to associate negative features to the outgroup. The *predication* sub-strategy, on the other hand, can be realised with different linguistic elements (adjectives, prepositional phrases, relative clause, verbs nominalization that describe actions or events in particular ways) that ascribe particular qualities to people (Hart [2010]:66). Summarising, both sub-strategies concretely consist in adding new representations and strengthening, weakening, or delaying existing ones (Hart [2010]:67). Metaphors and topoi are often used to achieve these sub-strategies.

#### Metaphors

Lakoff and Johnson (1980) define the metaphor as an element that is intrinsic in language, as it is one of the bases of our cognitive activity. For this reason, metaphorical structures cannot be considered as arbitrary, but reflect strongly our physical and cultural experience of the society. Metaphors can represent actual transmission vehicles of the dominant ideology in the analysis of political and social discourses. Moreover, they are relevant for the creation and development of stereotypes (Teun A. [1993]). Saiz de Lobado and Bonomi (2012) investigate the use of metaphors and metonymies (a special kind of metaphor where an object is conceptualised through its relation with another object in the same domain) in the media discourse about immigration. They schematise a methodology to recognise conceptual metaphors and metonymies in relation to immigration in four steps (2012:119-121). We reduced it to the two steps that are going to be mostly relevant for us:

• Determine if the main linguistic elements related to immigration are used with their literal meaning, if it's not the case:

- For metaphors: determine if the contextual meaning differs from the literal one;
- For metonymy: identify the main relation between elements;

#### Topos

In rhetoric topos is generally defined as a recurring and traditional concept or theme. We will base our analysis on the topos-concept of Wengeler (2003), strongly connected with the argumentation theory. Topos is here intended as connected to the enthymeme on which it is based. Wengeler defines enthymeme as "rhetorical syllogism". While logical syllogisms need to be formal-logical processes, rhetorical ones just need to have an almost-logical conclusion, that has consequently to be just plausible and not real (Wengeler [2003b]:60). Enthymeme describes therefore the prototypical argumentation process that includes an argument, a rule of derivation, and a conclusion. The importance of the rule of derivation is not to be underestimated: this rule is in charge to legitimate the passage from an uncontroversial argument to a controversial conclusion.

(2.1) argument: the right of asylum is misused by many derivation rule: a right that is misused should be changed conclusion: the asylum right should be changed

Example 2.1 (Wengeler [2003b]:61) shows exactly how the derivation rule works, starting from real premises and leading to a completely unjustified conclusion. Enthymeme are often used in public discourses even without the necessity to express the rule of derivation and the conclusion. Wengeler distinguishes two kind of topoi, appealing to the differentiation already expressed by Aristotle of *general enthymeme* and *specific enthymeme*. *General enthymeme* build upon identical structures and can be recognised in every discourse. They often comprehend the more-less, part-totality topoi, etc. *Specific topoi*, called "context related rule of derivation" by Wengeler (2003:62), express general thoughts or expressions typical for a specific discourse. In our investigation we will be mostly interested in *specific topoi*, in particular in the causal ones. These are topoi showing a causal pattern structure, meaning that one fact implicates directly and automatically another fact.

- (2.2) Topos of overload: since one person/institution/country is or risk to be overloaded with certain problems, a determinate action should be taken in order to reduce or prevent the load.
- (2.3) Topos of economical benefit: since an action produces or does not produce an economical benefit, it should be actuated or not actuated.

As shown in the two examples 2.2 and 2.3, argument and conclusion are strongly context related, while the derivation rule, which can be realised in various linguistic ways, is the same, consisting of a causal relation.

#### 2.2.2.1 Ingroup vs. Outgroup

In the discourse about migration, the strong differentiation between the ingroup and the outgroup constitutes the basis of every kind of coercion strategy (section 2.2). This dichotomy, constructed using specific linguistic strategies that we are nearby going to introduce, is already present at a social-legal level in our society.

Concepts like *migrant*, *immigrant*, and *migration*, all implicitly presuppose the concept of *citizen*, to which they are opposed (Guild [2009]). The concept of *citizen* plays nowadays a vital role, as it justifies the strategy adopted by modern states consisting in the employment of two different treatments for distinct individuals. *Citizenship* is, in fact, the principle which determines the possibility or the impossibility for an individual to access goods and services of a determinate state. Furthermore, Guildt affirms "the principle of equality among citizens, famously arising from the French revolution, has become also the dividing line between entitlement and exclusion". Individuals who do not posses the legal and social status of citizen are foreigners, the *foreigner status* is sufficient to both their negation of safety and benefits, and their territorial exclusion.

This social distinction is of course reflected in the discourse about migration, and it provides the basis for the self and the other representation. As Blackledge says (2005:21), in the self and other representation both strategies, referential and predicational are used. Referential strategies consist often in metaphorical or metonymical processes that represent people in terms of specific characteristics, which they share with others. These characteristics that often implicate negative evaluation are chosen to represent a whole group. Individual parts of the group can be referred in terms of national, linguistic, ethnic, cultural, economic, legal, religious, or sexual characteristics. Predicational strategies, on the other hand, are employed for assigning qualities to persons, objects, and events that are linguistically more or less evaluative. Traits such as *lazy*, greedy, dirty, diseased, refusing to integrate, abusive, oppressive to their women, criminal, culturally primitive, responsible for unemployment are often associated and predicated to immigrant groups. We will now introduce the three strategies of deixis, referential metonymy, and positive and negative face theory that can all be used to generate and reinforce the ingroup outgroup dichotomy.

#### <u>Deixis</u>

In the representation of this dichotomy, deixis plays obviously a determinant role. It consists in the meaning of a linguistic expression whose interpretation is relative to the extra linguistic context of the utterance (expression). It can include information such as the individual who is speaking and the place or time of the speaking: the three coordinates articulated around the communicative event (Antelmi [2012]:87f.). Chilton (2004:56) highlights the importance of personal deixis in the analysis of the political discourse, as especially pronouns can be used to define specific group identities. Spatial deixis is also important since it performs a relation to a geographical or political space, positioning verbs like *come/go*, *bring/take* in conventional frames. Deixical centre (me-here-now) is sometimes used to underline the implied "anchoring point" that text producers and text consumers impose or construct during any communicative interaction. This anchoring depends upon political discourses about cognitive frames that embody conventional shared understanding of society's structure.

#### Referential Metonymy

Saiz de Lobado and Bonomi (2012:127-130) identify recurring metonymies related to the term used to define the actors of migration. They define metonymy as a particular kind of metaphorical process that consists of a conceptualisation of an object through its relation with another object. It differs from the metaphor because it consists in a referential process that is realised in the same conceptual domain. Metonymy is characterised from a mostly automated strong relation between two elements making part of the same domain. They consist in relational processes like cause-effect, part-whole etc. The two researchers identify three typical kinds of referential metonymy for the discourse about immigration, which namely imply reference to the subject based on:

- Legal status (i.e. illegal immigrant);
- Geographical origin (i.e. Syrian);
- Emotive status (i.e. desperate);

From an argumentative and pragmatic point of view these references have evident effects. The legal-status reference creates a permanent bounding between the individual and its "absent" legal condition. This contributes to the establishment of negative association and prejudice connected to the individual. The geographical reference does not directly create negative association, but it often consists in superfluous information, not necessary to the codification of the message. The frequent use of these reference methods in sentences describing negative events or problems leads to a mental association in the text consumers between the geographical origin of the individual and negative events. Finally, the third kind of referential metonymical process, consisting of the individual's emotive status, which is often a negative one, also contribute to create negative association. Individuals are not perceived as humans able to feel positive and negative feelings but are only associated with negative ones.

#### Positive and Negative Face Theory

Another strategy, often used in the contrastive representation of the ingroup and the outgroup relies on the *face* concept. The face theory is strongly dependent on concepts like *respect* and *politeness*. In fact, every speech act can be perceived as polite or impolite depending firstly on the utterance, and secondly on the social role of the participants. Goffman (1967) introduces for this purpose the notion of *face*. Goffman defines *face* as " the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact" (Goffman [1989]:5).

Brown and Levinson (1987) conceptualised the first prominent model - that claims to be universal - starting from Goffman's face concept. They introduced the distinction between *positive* and *negative* face. Positive face reflects the need to be appreciated and accepted as part of a group, while negative face reflects the need to have independence and freedom in thoughts and actions. This theory considers the necessity for people in communication to balance the two contrastive feelings: the positive need to be accepted as an insider and to be part of a group, as well as the negative need to maintain their personal action's freedom. Chilton (2004:39f.) applies the face theory to politics. He assumes the addressing positive face by expressions which appeal to *patriotism*, *brotherhood*, *civilised values*, and similar concepts. Basically, all those concepts that have as part of their frame some notion of the positive characterisation of the ingroup. On the other hand, the process of addressing negative face consists in a minimisation of the possible dangers to the freedom and security of the ingroup individuals and of the community that they constitute. The typical verbal behaviour to fulfil this addressing is, according to Chilton, referring to threatening referents not directly, but obliquely or by euphemism.

#### 2.2.2.2 Emotive Coercion

Modern cognitive science has exceeded the traditional division between emotion and reason. These two entities are seen as integral parts of cognitive processing. In fact, emotions are very important for every individual and play a large role in decisionmaking (Antelmi [2012]:215). For our discourse, we are mostly interested in negative emotions raised by the use of topoi and metaphors when talking about immigration.

Anger and fear are the most considered emotions in the discourse about immigration (Hart [2010]:80). Fear is a kind of emotion that responds to signals caused by physical threats, more than to social ones. Anger, on the other hand, has been seen as a defence mechanism that implements the cheater-punishment strategy. This emotion is deeply rooted in social altruism (Pinker [2002]:272). Anger, according to Pinker, is a defence mechanism that protects people whose altruistic behaviour is being exploited. These angry people will interrupt the relationship with the exploiter, thus avoiding future altruistic behaviours to be exploited again. Evoking emotions is an important technique that can be employed in strategic discourses. It lead text-consumers to construct cognitive structures that associate the outgroup with social or physical threats. Hart groupes these strategies under the umbrella of emotive coercion. As concrete implementations, we can identify the use of determinate *metaphors* and the *topoi of danger, displacement, and disease*.

Saiz de Lobado and Bonomi (2012:122-126) analyse some typical metaphors identifying three particular kinds:

- Immigrants as natural disaster;
- Immigrants as conflicts or war;
- Immigrants described using biblical references;

One of the most frequent metaphors is the one that associates the image of a natural disaster to immigrants. They are, in fact, often compared to *waves*, *flow*, or *surges*. These images always tend to identify immigrants as a mass and never as individuals. Furthermore, the association with natural disaster that cannot be stemmed, or handled, contributes to the establishment of fear and negative association connected to immigrants.

The second kind of metaphor identified by Saiz de Lobado and Bonomi is the one associating immigrants with conflicts or war. Immigrants are perceived as enemy forces that threaten the national borders which need to be defended. In the same way, arrivals and landings are understood as invasions to fight. Fear and anger are caused by this metaphorical field.

Finally, the last kind of metaphor is the one associating immigrant with biblical elements. This case presents less variety in the expressions related to religious concepts. It shows, however, significant recurring linguistic elements like in the case of *exodus*. *Exodus* used to describe immigration is connected to a negative stereotype that causes fear. Indeed, it refers to a mass of individuals without destination, hunted from their nations and in search for a new place to resettle.

#### Topoi of danger, displacement, and disease

Hart presents three typical topoi used in the immigration context to evoke anger or fear (2010:82-83).

In the topos of danger immigration is presented as dangerous for the ingroup members. Danger is often connected with the metaphor of naturalisation or militarisation. In the first metaphorical type immigration represents a natural calamity itself (flood, wave), in the second one it is seen as an attack, an invasion, or an aggression.

This topos often co-occurs with the topos of displacement, where the danger becomes the displacement of the ingroup. Most frequently, forms of this topos represent the outgroup becoming the majority and gaining control aver the ingroup, thus dictating new cultural values and laws. Therefore, this topos poses a threat to the cultural identity of the ingroup.

Topos of danger also sometimes co-occurs with the topos of diseases: xenophobic propaganda often tends to associate immigrant to bearer of diseases that could possibly infect the ingroup members.

All these topoi lead to the innate fear of:

- Physical harm;
- Loss of territorial control and cultural identity;
- Infection from transmittable diseases;

#### 2.2.2.3 Cognitive Coercion

Cognitive coercion consists, as previously mentioned, in a form of coercion performed by the activation of specific cognitive processes. One of these, identified by Hart, is the activation of the cheater-detection module (2010:71). Every society is based on numerous contracts that regulate the position of every individual in the society. Under social contract we understand the fact that if an individual uses society resources he should somehow contribute to it and to its effectiveness. Hart shows how topoi constructing a cognitive association according to which immigrants violate the social contract are likely to activate the cheater-detection modus of text consumers. In this process the text consumer identifies itself with the society, as part of the ingroup being exploited by the outgroup. This leads to anger, which is directly related with anti-immigrants feelings and actions. A specific lexicon contributes to the activation of the cheater-detection modus. This lexicon encompasses antonymic verbs pair like *give/take*, abstract nouns like *kindness*, *generosity*, and words in the semantic field of compensation like in example *contribute*, *give back*, *repay*, *in return*. As strategies to activate this modus, Hart identifies the *topoi of disadvantage*, *burden finance*, and *displacement*, the *topos of exploitation*, and the *topoi of culture* and *crime* (Hart [2010]:72).

#### Topoi of disadvantage, burden finance, displacement

The topos of disadvantage consists in the statement that immigrants do not bring economical advantage to the ingroup. They are believed to be lacking of any useful attribute that could contribute to the ingroup welfare. Lexical units often connected to this topos are *uneducated*, *unqualified unskilled*.

The topos of burden finance consists in reporting immigrants in monetary terms as directly costing a specific price to the ingroup members.

Finally, the topos of displacement describes members of the outgroup having access to special socio-economic resources - that are possibly negated to the ingroup members - at expense of the ingroup.

#### Topos of exploitation

In this topos immigrants are represented as social cheats. They become freeriders that take from the ingroup without giving anything back in return. Relying on group resources but not engaging in collective action is presented as an exploitation of the ingroups generosity. This topos is also often realised through metaphorical language, picturing, for example, immigrants as organisms that exist in parasitic relation rather than mutualistic symbiosis with their host. Exploitation is also understood in relation to the system itself: for example in the idea of immigrants abusing the asylum system. The typical differentiation between genuine asylum-seekers and economic migrants is a perfect example of this phenomenon. This referential strategy simultaneously serves a predicational function. In fact it profiles profit as opposed to persecution as the motivation for migration. This is often used as premise in order to change the current law (i.e. asylum law).

#### Topoi of culture and crime

The topos of culture is based on the argumentation that because the culture of a specific group of people is as it is, unsolvable problems will arise. The outgroup is

presented to have different values and culture to the ingroup and being unwilling to assimilate. Outgroup members are consequently defined on the basis of these predicated differences, realising also a referential strategy. Language is often used to underline the cultural difference. This topos supports the topos of disadvantage that we introduced before: since immigrants are not willing to integrate they are definitely not going to contribute to the society.

The topos of crime consists in describing the immigrant as a criminal. The referential expression *illegal immigrants* already contributes to this phenomenon. This topos leads to the establishment of the exploitation one.

# **3 Corpus Linguistics**

In this chapter we will introduce *Corpus Linguistics (CL)*, the second research field of the theoretical background of this investigation. After briefly discussing the historical development of the discipline, we will discuss in detail the extraction methodologies that we are going to use in a further step.

CL is a method for carrying out computer-aided linguistic analysis on corpora, which are defined as stored collections of samples of naturally occurring language. This approach can be used to describe different kinds of linguistic features and to test hypotheses formulated in various linguistic frameworks. Its strength lay, among other things, on the possibility to test linguistic hypotheses relying on large amount of authentic spoken or written language. This removes the researcher from having to rely on its own intuition, which cannot always be objective.

CL, considered as a sub-discipline of linguistics, is datable to the post-war period, around the late '50s. Computer technology had to develop systems to manipulate large amounts of computerised text before this sub-discipline could possibly emerge (McEnery and Hardie [2013]:728). CL was developed especially in opposition to Noam Chomsky, who favoured a rationalist approach rather than the empiricism associated with the corpus-based approach like McEnery and Hardie notice (2013:729 - 732). Determinant to the development of the discipline was the English Corpus Linguistics since "the major, systematic contributions of corpus linguistics to the improved descriptions of the lexis and grammar of language were made within [it]". (2013:732). Finally, the creation of the Brown Corpus which was "the first of the modern, computer readable, general corpora"<sup>1</sup>, edited in 1967 by Henry Kucera and W. Nelson Francis at the Brown University in Rhode Island, led to the progressive development of many other corpora. The Brown Corpus contains 500 samples of English-language text with around one million tokens. It was not only used in many CL studies, but it also set the parameters for the sample corpus. Sample corpora aim to represent a specific kind of language variety in a determinate point in time (McEnery and Hardie [2013]:733).

<sup>&</sup>lt;sup>1</sup>http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus\_ling/content/ corpora/list/private/brown/brown.html

Nowadays, corpora are used for scientific researches and observations in different disciplines such as linguistics, history, sociology, literature, and political science. As we have seen, the concept of corpus is not new, but its meaning changed in the last 25 years. Corpora were, in fact, directly influenced by the advancement of the digital revolution (Mayaffre and Poudat [2013]:65). The most obvious impact consists in the easiness in building corpora in terms of resources' availability. More and more documents are being digitalized in every research field, at the same time contemporary documents are often directly made available on the web. Furthermore, different open source tools explicitly created to generate corpora from online pre-existing text are disposable.

## 3.1 Corpus-Driven and Corpus-Based

As Partington shows, the co-penetration of DA and CL has not always been perpetuated (2004:1). This was due to the fact that many early corpora contained just fragments of texts rather than whole ones. Furthermore, even when larger and more complete corpora were available did CL mostly develop methods to find evidences of linguistic phenomena fast and easily. This approach, called *corpus-based* is obviously not optimal for analysing discourses in a Foucaultian way (Bubenhofer [2008]:411). In fact, the *corpus-based* approach necessitates pre-formulated hypothesis that are not compatible with the main ideas of DA. However, this is not the only possible approach. As Bubenhofer points out a changing of perspective leads to another approach: the *corpus-driven* one (2008:412). The corpus is here understood as an amount of data that shows determinate structures. Following Taubert:

While corpus linguistics may make use of the categories of traditional linguistics, it does not take them for granted. It is the discourse itself, and not a language external taxonomy of linguistic entities, which will have to provide the categories and classifications that are needed to answer a given research question. This is the corpus-driven approach. (2005:4)

As Teubert affirms, the *corpus-driven* approach consists in the try to understand the corpus not as a "reference book" but rather as an amount of data that shows structures classifiable only at a later time. In fact, to perform DA we need to identify the patterns and argumentative strategies of section 2.2 without already knowing which lexical elements or morpho-syntactical structures they consist of. In the two next sections we are going to provide a way to methodologically solve this task. We are going to introduce the extraction of keywords, co-occurrences, and multi-words-expressions (MWEs), in order to perform a *corpus-driven* approach. As Bubenhofer points out (2008:421), with the purpose of performing a meaningful analysis it is necessary to start from the corpus-driven extracted significant elements and then proceed corpus-based with the traditional analysis methods of CL: verification of the distribution and categorisation of the results. In our case, we are going to analyse the extracted significant elements in their context and categorise them into the frame of the analysis categories discussed in section (2.2). The ideal process to perform DA consists in a continuous transition between the corpusdriven and the corpus-based approach. This alternating between the corpus-driven extraction of data and its corpus-based analysis, with the constant addiction of new corpus-driven extraction material, leads, to a more and more accurate discourse description and analysis (Bubenhofer [2008]:422).

## 3.2 Extraction of Significant Elements

In this section we are going to introduce three processes vital for a corpus-driven approach: keywords extraction (section 3.2.1), co-occurrences extraction (section 3.2.2), and MWEs extraction (section 3.2.2).

As Seretan affirms (2011:31), the general procedure of every extraction consists of two main steps:

- 1. Candidate identification through the use of specific criteria;
- 2. Candidate ranking with a given critical value;

Before performing the first step, a linguistic preprocessing step - especially for languages with rich morphology and free word order - is recommended. The procedures often performed are Part-Of-Speech tagging, lemmatisation<sup>2</sup>, and parsing. POS-tagging consists in the automatic application of grammatical category (part of speech category) to every token in a text. Lemmatisation, instead, consists of merging all the inflected word forms to their lemma. The term *lemma* defines the base of a word without its possible inflection. Every headword found on the dictionary is a lemma. Finally, parsing is a process that breaks down a text in its syntactical components. It is usually performed to disambiguate ambiguous sentences by identifying the dependencies between their elements. To perform these processes, linguistic analysis tools such as Part-Of-Speech-Taggers or Parsers can be used (Seretan [2011]:31).

<sup>&</sup>lt;sup>2</sup>It has been stated that lemmatisation brings advantages in this process, but this opinion is still debated in this field of research. To deepen this question we refer to Bubenhofer [2009]:124 and Tognini-Bonelli [2001]:92.

#### 3.2.1 Keywords Extraction

The first methodological operation to perform in a corpus-driven approach is to extract keywords. Keywords are here intended as words or concepts of great significance that are typical for a determinate discourse. In this context, typical should not be confused with frequent but rather defined as "frequent in regards to a specific portion of language" (Bubenhofer [2008]:409). To pose the problem in other terms: we are searching for the words that best characterise the differences between two corpora (Kilgarrif [2001]233).

In order to identify the keywords of a specific corpus contrastively to another corpus we need to confront the observed values with the expected ones. Observed values are the values actually calculated from the corpus. Expected values consist, on the other hand, in the values that we would expect if the two corpora were equally balanced in their occurrences of keywords. To compare them, contingency tables such as Table 1 are needed. Contingency tables are usually used in statistics to display sample values in relation to two or more random variables that can possibly be dependent on each other (Seretan [2011]: 32). The Null-hypothesis consists in affirming the independence of two random values. In this method, a word X is recognised as keyword if the Null-hypothesis is rejected. This is done by proving the possible relevant difference in the occurrences of word X in the two corpora.

	Corpus 1	Corpus 2	Total
Frequency of word X	А	В	A + B
All other words	С	D	C + D
Total	A+C	B+D	A+B+C+D

Table 1: General Contingency Table

Following Bubenhofer<sup>3</sup>, the first step is to calculate a contingency table of the observed values. For this we need to compute the occurrences of a hypothetical word X, but also of all the other words that are not X for each Corpus (Corpus 1, Corpus 2). In fact, following Table 1 Corpus 1 contains A, all the occurrences of word X, and C, all the occurrences of the other words. Corpus 2, in the same way, contains B, all the occurrences of word X, and D, all the occurrences of the other words.

<sup>&</sup>lt;sup>3</sup>Induktive Statistik: Signifikanztest 1: http://www.bubenhofer.com/korpuslinguistik/kurs/ index.php?id=statistik\_signifikanzChi.html

To fill in the contingency table with the expected values a simple Rule of Three calculation is needed. This aims to compute the expected values in relation to the size of the two corpora.

$$\frac{a}{b} = \frac{c}{x} \longrightarrow x = \frac{bc}{a} \tag{3.1}$$

As show in the formula 3.1, the expected values - depending on the size ratio of both corpora - can be computed with the proportion calculation. A number of significance tests can be applied once that both contingency tables have been built.

One standard test for identifying significant words is the *Chi-Squared test* (also called  $X^2$  test)

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
(3.2)

In the formula of example 3.2 the variable O represents the observed value and the variable E the expected one. To obtain  $X^2$  we need to calculate  $(O - E)^2/E$ for every cell value of Table 1 and then sum all the results together. The obtained number corresponds to our final value for word X (Bubenhofer [2009]:137).

To evaluate the obtained value, statistics provides the so-called *Table of Chi*square. This table contains critical values for X depending on degrees of freedom (df) and probability (P).

df	P = 0.05	P = 0.01	P = 0.001
1	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.35	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52

Figure 1: X Critical Value

The degree of freedom is directly co-related to the size of the table. It is computed with:

$$(NumberOfRow - 1) * (NumberOfColumn - 1)$$

$$(3.3)$$

In the case of a contingency tables with size 4 \* 4 - like Table 1 - the calculation is:

$$(2-1) * (2-1) = 1 \tag{3.4}$$

Consequently, the relevant degree of freedom is 1, implying that the relevant row of Figure  $1^4$  is exclusively the first row. The other df values become relevant with bigger tables. P defines, on the other hand, the degree of probability. It determinates the probability according to which the Null-hypothesis can be rejected. In our case we have probabilities of 0,05%, 0,10%, and 0,001%.

Summarising, the first column of the Chi-square Table determinates the minimum value that  $X^2$  needs to reach so that the frequency difference between the two corpora can be considered as 95% significant. In other words, if  $X^2$  value for word X is greater than 10.83, there is a 99.9% probability that the word is a keyword.

#### 3.2.2 Co-occurrences and MWEs Extraction

Two other statistical methods in a corpus-driven approach consist in the extraction of co-occurrences and of MWEs.

Co-occurrences are defined as two adjacent words that statistically co-occur contrastively to other pairs of words that could possibly co-occur in a corpus (Seretan [2011]:29). The MWEs concept has been defined in many ways. Traditional definitions mainly focus on their idiosyncrasy at lexical, syntactic, semantic, pragmatic, and/or statistical level (Calzolari et al. [2002], Kim [2008]). One simple English example of MWE is:

#### (3.5) Kick the bucket

To understand the expression the text consumer needs to over-goes the words boundaries. The expression at example 3.2.2 means to die rather than to physically hit a real bucket.

The extraction procedure for co-occurrences is similar to the one explained in section 3.2.1 but it presents some slight differences. The extraction is not performed contrastively between two corpora but rather between the two words that hypothetically co-occur. The main difference is therefore in the elements that fill the contingency table. As Seretan explains, to detect a dependency between the components of a candidate pair, we can introduce two discrete random variables (X

 $<sup>{}^{4}</sup> Table \ of \ Chi-square \ statistics: \ {\tt http://alexei.nfshost.com/PopEcol/tables/chisq.html}$ 

and Y). X variable will be associated with the first position of the pair and Y with the second one (2011:33). This allows to fill in the table, considering the occurrence of one particular pair X contrastively to the occurrence of all the other pairs in the corpus.

	Y = v	Y = -v	Total
X = u	А	В	A + B
X = -u	С	D	C + D
Total	A+C	B+D	A+B+C+D

Table 2: Contingency Table for Co-occurrences

In this case as shown in Table 2:

- Variable A represents the occurrences of the pair X in the corpus (joint frequency of u and v);
- Variable B represents the number of pairs that have u in the first position and -v in the second one;
- Variable C represents the number of pairs that have -u in the first position and v in the second one;
- Variable D represents the number of pairs with -u in the first position and -v in the second position;

Consequently, A + B consists of the frequency of all pairs that have u in the first position, and C + D indicates the number of items that have -u in the first position. In the same way, A + C indicates the frequency of all the pairs with v in the second position, and B + D refers to the frequency of all the pairs having -v in the second position. These numbers are referred to as marginal frequencies. A + B + C + D corresponds to the total number of candidate pairs that are present in the corpus. The variable D is not so trivial to compute in practice, since we normally compute occurrences and not non-occurrences. However, it can easily be calculated with the simple equation of example 3.6.

$$D = N - (A + B) - (A + C) + A$$
(3.6)

Once the contingency table is filled, we can proceed with the significance calculation. Another kind of calculation suitable for this purpose is the Log-Likelihood-Test. It can be accomplished using the following formula:

$$G^{2} = 2[AlogA + BlogB + ClogC + DlogD - (A + B)log(A + B) - (A + C)log(A + C) - (B + D)log(B + D) - (C + D)log(C + D) + (A + B + C + D)log(A + B + C + D)] (3.7)$$

This computation performs the association score for a candidate pair X (u,v) by executing the contrast of two likelihoods and considering the logarithm of the result (Seretan [2011]:39):

- The likelihood of the observed counts in the contingency table under the Nullhypothesis (hypothesis of independence);
- The likelihood of the observed counts in the contingency table under the alternative hypothesis (hypothesis of dependence);

The Log-likelihood<sup>5</sup> ratio expresses, therefore, the occurrence of a pair X in a corpus under the Null-hypothesis model and the alternative model. The obtained value is then compared to the critical value of the Chi-Squared test of Figure 1 to decide whether the Null-hypothesis has to be rejected or not, as explained in section 3.2.1.

The extraction of MWEs, on the contrary, is more similar to keyword extraction. It is contrastive between two corpora and it requires two contingency tables, one with observed values and the other with expected values. To calculate the significance of MWEs, Chi-Squared algorithm and Log-likelihood can be used. The only difference in respect to the keywords extraction procedure consists in the usage of n-grams instead of single words. Once that the chosen n-grams are computed and extracted in a data structure, the procedure, as we said, is exactly the same as the other one.

<sup>&</sup>lt;sup>5</sup>For a more detailed explanation of the Log-Likelihood Formula we refer to works like *Syntax-Based Collocation Extraction* by Seretan (Seretan [2011]) or *Corpora and Collocations* by Evert (Evert [2008]).

# **4** Statistical Machine Translation

In this section we will furnish an introduction to Statistical Machine Translation (SMT), focussing mainly on the training on the Translation Model (TM).

SMT consists in the usage of statistical methods to produce translations from a source into a target language. SMT systems consist mainly of two models:

- Translation Model (TM);
- Language Model (LM);

The TM is the statistical model that produces collections of possible translations and their probabilities. To train this model it is necessary to use large amounts<sup>1</sup> of human-translated parallel corpora (Carstensen et al. [2010]:647f.). The parallel corpus has to be sentence aligned in order to know which sentence of the source text corresponds to which sentence of the target text. Sentence-alignment is usually performed comparing different elements, such as the lengths of sentences and anchor points, in both texts. The latter consist in tokens that are identical or very similar in source and target languages (i.e. numbers or name entities). The one-to-one sentence alignment is the standard alignment, although different alignments, such as many-to-one and one-to-many alignments are as well possible (Carstensen et al. [2010]:648). From the sentence aligned corpora the TM calculates word alignment. Given a pair of translated sentences the model has to determinate which word w1 in sentence S1 is a translation of which word w2 in sentence 2. In thi case too, many-to-one word alignments are possible. furthermore, the same word order in both sentences cannot be assumed.

The TM is based on IMB models that apply the Expectation Maximisation algorithm. The word-based model works by translating each word independently. Every translation probability is then used to calculate the translation probability of the entire sentence (Koehn [2010]:86f.). To translate individual words and to calculate their corresponding translation probabilities two operations are needed: word alignment and lexical translation probability distribution. These two calculations

 $<sup>^1\</sup>mathrm{To}$  generate an acceptable model at least 1. Million words are needed.
are inter-dependent: the translation distribution is needed to compute the word alignment, and the word alignment is needed to calculate the translation distribution. To solve this issue the Expectation Maximisation algorithm is used. In fact, in a first step the model is initialized with a uniform distribution. In the expectation step this initial distribution is applied to the data. The result of this step is that words from source and target language are aligned. In the following maximization step the translation probabilities are calculated based on the expected alignments. Afterwords, the alignment is estimated again, this time based on the calculated lexical translation probabilities. These steps are iterated a number of times, until convergence is reached (Koehn [2010]:88). Various outputs are obtained from this first model: the most relevant for us are the lex-files, containing the calculated probabilities for each pair of words in both directions (source to target language and target to source language).

Nowadays, the most used model is the phrase-based one. This model differs from the word-based one because it calculates the translation probability of phrases<sup>2</sup> - sequences of words - that can differ in lengths. The phrase pairs are saved in phrase tables with the corresponding probability (Koehn [2010]:128-131). After the translation the phrases need to be reordered since the order often differs in source and target language. To perform this reordering a distance-based reordering model is used (Koehn [2010]: 127-129).

The second step of the translation system consist in the LM of the target language. To train this model a corpus in the target language is necessary. Both the text in the target language of the training corpus and an additional monolingual corpus can be used as corpus The LM aims to generate a fluent translation in the target language, therefore, it contains the language's typical word groups and word order (Carstensen et al. [2010]:650). LM are based on the calculation of probabilities of n-grams, which are the likelihood that certain words follow each other. Since in our investigation we are going to train a simple word-based TM, we are not going to explain the LM procedure in further detail<sup>3</sup>.

<sup>&</sup>lt;sup>2</sup>Phrases are not to be intended in a strictly linguistic way, rather as general multi-words-units. <sup>3</sup>For more information about the procedure consult Koehn [2010]:181f.

## 5 Corpora

In this investigation we used two kinds of corpora: primary and translation. Primary corpora - consisting of political press releases - served as basis to extract the actual investigation material such as keywords and other significant elements (see section 3.2). Translation corpora - German-Italian parallel corpora - were instead used to train a simple SMT system which allowed us to compare the discourses (see section 4). In this section we will present the corpora introducing their characteristics.

## 5.1 Primary Political Corpora

We decided to build two specific corpora containing the political discourse of both Italy and Switzerland. In the field of political linguistics numerous classifications of political text typologies have been proposed. Following Dell'Anna (Dell'Anna [2010]:37) we distinguish between primary and secondary political texts, using the pragmatic criterion of the text consumer. Primary political texts are defined as produced by politician in relation to electoral or institutional facts. On the contrary, secondary political texts can be produced also by non-politician and arise from the observation or interpretation of political facts. We decided to focus only on primary political texts, limiting the contents of the corpora to press releases<sup>1</sup> published on the web pages of the main political parties in both countries. This was done under the assumption that primary political texts represent the actual political discourse since these refer to the respective electorate without any mediation, as it is the case for secondary political ones.

<sup>&</sup>lt;sup>1</sup>We extracted articles from the sections *Medienmitteilungen* and *Medienkonferenzen* for the Swiss-German corpus, and *Comunicati Stampa* and *Rassegna Stampa* for the Italian Corpus. For the parties without this differentiation, we extracted articles from the *News* section.

### 5.1.1 Corpora Construction

We extracted the articles with the web data extraction platform import.io. Table 3 gives the overview of the corpora composition.

C.								Tot
СН	SP 09.01.12 14.01.16 1066 A.	SVP 14.04.08 21.12.15 598 A.	CVP 05.01.09 08.01.16 1751 A.	Grüne 10.03.01 12.01.16 1518 A.	EVP 12.02.05 14.06.12 420 A.	GLP	BDP	7'231 A.
ITA	PD 23.02.15 09.01.16 907 A.	PDL 21.01.13 08.01.16 1982 A.	SC 11.01.13 12.01.16 761 A.	SEL 16.01.14 09.01.16 3024 A.	LN 14.01.14 12.01.16 3014 A.	CeR 18.05.15 09.01.16: 68 A.	FI 01.01.13 02.12.15 103 A.	12'289 A.

### Table 3: Corpus Information

Both corpora consist of a different number of articles, which cover a different time sequence, depending on the availability of the various parties. Despite the differences in size, quantity, and time overlapping between the articles, we consider the corpus as being representative for the general political discourse of both nations.

The Swiss-German corpus consists of articles extracted from different ideological oriented parties<sup>2</sup>. We exclusively extracted articles on the German version of the parties site. Table 4 shows an overview of the parties that compose the Swiss-German corpus.

<b>Right-Wing Parties</b>	<b>Center-Wing Parties</b>	Left-Wing Parties
Schweizerische Volkspartei (SVP): Right- populist and National- conservative	Christlichdemokratische Volkspartei der Schweiz (CVP): Bourgeois and Christian-Democratic	<i>Sozialdemokratische Partei der Schweiz (SP):</i> Social-Democratic
<i>Freisinnig-Demokratische</i> <i>Partei (FDP</i> ): Bourgeois and Liberal	<i>Evangelische Volkspartei</i> <i>(EVP)</i> : Evangelic and Conservative	<i>Grüne Partei der Schweiz</i> ( <i>GPS</i> ): Ecological, Pacifist and Feminist

### Table 4: Swiss-German Parties

We decided to exclude the *Bürgerlich-Demokratische Partei (BDP)* and the *Grünliberale Partei (GLP)*, two minor centre-wing parties, because of lacking of available online material. The Swiss-German articles are dated between January 2006 and January 2016. The complete Swiss-German corpus consists of 7'231 articles with around 2'739'000 tokens.

<sup>&</sup>lt;sup>2</sup>https://www.ch.ch/de/wahlen2015/parteien/

For the Italian corpus we proceeded in the same manner  $^3$ . Table 5 shows an overview of the parties that compose the Italian corpus.

Right/Center-Right-Wing parties	Center-Wing Parties	Left/Center-Left-Wing parties
<i>Popolo della liberità (PDL)</i> : Christian-Democratic and Liberal-Conservative	<i>Scelta Civica (SC</i> : Liberal and Pro-Europe	PartitoDemocratico(PD):Social-Democratic,Christian-DemocraticandPro-Europe
<i>Conservatori e Riformisti (CeR</i> : Conservative and Liberal	<i>Unione di Centro (UDC)</i> : Christian-Democratic	<i>Sinistra Ecologia e Libertà</i> <i>(SEL)</i> : Social-Democratic and Eco-Social
<i>Fratelli d'Italia - Alleanza Nazionale (FI)</i> : National and Conservative		
Lega Nord (LN): Federalist and Right-Populist		

#### Table 5: Italian Parties

The Italian articles are dated between January 2013 and January 2016. The complete Italian corpus consists of 12'289 articles with around 3'220'000 tokens.

We consciously excluded the recent 5 Star Movement (Movimento 5 Stelle) from the corpus despite his contemporary representativeness in the Italian politics. This choice is based on the political status of the party that defines itself as a "libera associazione di cittadini"<sup>4</sup> contrasting form and methods of traditional parties. The 5 Star Movement, moreover, officially rejects the public media channel, considering the media system part of the corrupted elite. For this reason all the communications between party and voters take place on the movement's blog, Il Blog di Beppe Grillo<sup>5</sup>, an influential medium that connects a network of citizens. This consequently affects the communication and press releases system of the movement, making it not suitable for our purpose, which aims to build a corpus as well balanced as possible.

### 5.1.2 Immigration Sub-corpora

In order to extract the keywords (see section 3.2.1) of the political discourse about immigration of both countries, and perform the first step of our investigation we needed to create two immigration sub-corpora. These should consist exclusively of

<sup>&</sup>lt;sup>3</sup>http://www.parlamento.it/1063

<sup>&</sup>lt;sup>4</sup>Beppegrillo.it - Chi siamo: http://www.beppegrillo.it/listeciviche/liste/fabriano/ chi-siamo.html

<sup>&</sup>lt;sup>5</sup>http://www.beppegrillo.it

the articles from the primary political corpora concerning the subject of immigration. In order to extract these articles we used a lexical approach, thus identifying with regular expressions words or character patterns that are part of the semantic field of immigration.

For the Swiss-German sub-corpus we matched words like Asyl, Migration, Einwanderung, Migrant, Immigrant, eingewandert, immigriert, Zuwanderung, and Zuzug, including their versions with lower-cased initial characters and all the words including the matched ones. These words were chosen as representative for the subject, under the assumption that at least one of them would appear in every article with immigration as topic. In the following code snippet we report regular expressions (regex) that we used to extract the Swiss-German articles about immigration.

```
for line in file:
```

```
feature1 = re.search('(F|f)l(\w)chtl',line)
feature2 = re.search('(A|a)syl',line)
feature3 = re.search('(A|a)syl',line)
feature4 = re.search('(E|e)inwanderung',line)
feature5 = re.search('(A|a)usl(\w^a)nd',line)
feature6 = re.search('(A|a)usl(\w^a)nd',line)
feature7 = re.search('(I|i)mmigrant',line)
feature8 = re.search('eingewandert',line)
feature9 = re.search('immigriert',line)
feature10 = re.search('(Z|z)uwanderung',line)
feature11 = re.search('(Z|z)uzug',line)
```

We initially also included the pattern Ausl to match all the words of the semantic field of Ausland/Ausländer but had to substitute it with Auslä to avoid the matching of Ausland. In fact, we noticed that the presence Ausland increased remarkably the number of the matched articles, including wrong data. Many articles containing this word talked about *import*, *export*, and *economical relationship with foreign country*, focusing on economy instead of immigration.

For the Italian corpus we proceeded in a similar manner, matching articles with words and expressions like *rifugiato*, *profugo*, *immigrato*, *asilo politico*, *ius soli*, *richiedente asilo*, *migrante*, *flusso migratorio*, *permesso di soggiorno*, and *diritto di asilo*, including capitalised initials and their feminine and plural conjugations, as shown in the following code snippet.

```
for line in file:
```

```
feature1 = re.search('(R|r)ifugiat', line)
feature2 = re.search('(P|p)rofug', line)
feature3 = re.search('(I|i)mmigra', line)
feature4 = re.search('(A|a)silo_politico', line)
feature5 = re.search('(P|p)ermess(i|o)_di_soggiorno', line)
feature6 = re.search('(R|r)ichiedente_asilo', line)
feature7 = re.search('(F|f)lusso_migratorio', line)
feature8 = re.search('(M|m)igrant', line)
feature9 = re.search('(D|d)iritto_(.+)?_asilo', line)
feature10 = re.search('(I|i)us_soli', line)
```

Even for Italian, we had to exclude the word *straniero* that matched expressions like *investitori stranieri*, *lingua straniera*, *capitali stranieri*, or *industria straniera*, contained in articles with subjects differing from *immigration*.

After both operations we ended up with immigration sub-corpora corresponding to around 10% of the total corpora:

- German immigration sub-corpus: 1036 articles, 429'950 tokens;
- Italian immigration sub-corpus: 1349 articles, 333'361 tokens;

## 5.2 Translation Corpora

The two parallel corpora were used to train a simple TM of an SMT system for the language pair German - Italian in order to semi-automatically confront the two political discourses. To train the model we used two different parallel corpora:  $Europarl^{6}$  and  $bulletin4corpus^{7}$ .

### 5.2.1 Europarl

The *Europarl Corpus* is a parallel corpus extracted from the web site of the European Parliament by Philipp Koehn (University of Edinburgh). It consists of "a collection of the proceedings of the European Parliament, dating back to 1996". (Koehn [2005]:1). It contains documents in 21 European languages: Romanic (French, Italian, Spanish, Portuguese), Germanic (English, Dutch, German, Danish, Swedish),

<sup>&</sup>lt;sup>6</sup>http://www.statmt.org/europarl/.

<sup>&</sup>lt;sup>7</sup>http://kitt.cl.uzh.ch/kitt/b4c/en/.

Slavik (Bulgarian, Czech, Polish, Slovak, Slovene), Finni-Ugric (Finnish, Hungarian, Estonian), Baltic (Latvian, Lithuanian), and Greek<sup>8</sup>. we used the release v7, published on the 15 May 2012. The parallel corpus German - Italian consists of around 1.8 million lines containing around 45 million tokens. The *Europal Parallel Corpus* was expressly created to perform statistical machine translations. It is therefore provided as two files one in German and one in Italian containing plain, sentences aligned text.

### 5.2.2 bulletin4corpus

The *bulletin4corpus* is a parallel Corpus created by Institute of Computational Linguistics of the University of Zürich<sup>9</sup>. It consists of a collection of news articles from the internet page of the Swiss bank Credit Suisse<sup>10</sup>. The Corpus is multilingual and available in 4 languages: English, French, German, and Italian. The size of the parallel corpus for the pair German-Italian is of 1448 articles per language, with around 1.70 million tokens. The articles in the corpus range from 2001 to 2015 and cover different theme sections like economy, banking, society, sport, and politics.

The *bulletin4corpus* consists of two files in XML format that contain for each word of each sentence word-index, lemma<sup>11</sup>, POS tag, and actual word form as shown in Figure 2.

```
<article n="a1" id="cs-2009-02-24-Crisis">
    <h1 cat="Europa">
        <s n="a1-s1">
            <w n="a1-s1-w1" lemma="Osteuropa" pos="NE">Osteuropa</w>
            <w n="a1-s1-w2" lemma="stehen" pos="VVFIN">steht</w>
            <w n="a1-s1-w2" lemma="stehen" pos="VVFIN">steht</w>
            <w n="a1-s1-w3" lemma="vor" pos="APPR">vor</w>
            <w n="a1-s1-w3" lemma="ein" pos="APPR">vor</w>
            <w n="a1-s1-w4" lemma="ein" pos="ART">einer</w>
            <w n="a1-s1-w5" lemma="schwer" pos="ADJA">schweren</w>
            <w n="a1-s1-w6" lemma="Kreditkrise" pos="NN">Kreditkrise</w>
            </w>
```

Figure 2: German bulletin4corpus

<sup>&</sup>lt;sup>8</sup>http://www.statmt.org/europarl/.

<sup>&</sup>lt;sup>9</sup>Release 02, 6. March 2016 by Institute of Computational Linguistics, University of Zurich. <sup>10</sup>https://www.credit-suisse.com/ch/en/news-and-expertise.html.

 $<sup>^{11}\</sup>mathrm{POS}$  Tag and lemmatisation of the *bulletin4corpus* were performed with the TreeTagger.

Furthermore, an alignment file in XML format is provided. It contains, as shown in Figure 3, the alignment type of each sentence in each article between the two language files. This file is determining to achieve a plain parallel corpus. The example of the Figure shows that sentence 1 of article 249 of language 1 (German) is aligned with sentence 1 of article 1 of language 2 (Italian), their alignment type is 1 to 1.

```
1 <?xml version='1.0' encoding='utf-8'?>
2 <linkGrp toDoc='CS_news_corpus_it.xml' fromDoc='CS_news_corpus_de.xml'>
3 <link type='1-1' xtargets=' a249-s1; a1-s1'/>
4 <link type='1-1' xtargets=' a249-s2; a1-s2'/>
```

Figure 3: bulletin4corpus Alignmentfile

# 6 Calculating Discourses

In this chapter we describe the application of different procedures used to fulfil a corpus-driven approach (see section 3.1) and a statistical discourse calculation. Large amount of data and its processability allow a quantitative approach through the usage of determinate algorithms to inductively detect the exemplariness of the data (Bubenhofer and Scharoth [2008]:2). For calculating both discourses by statistically determining their significant elements, we firstly extracted keywords of the two immigration discourses (see section 6.2). Afterwards, we used them to analyse their co-occurrence with other words in the immigration sub-corpora, working corpus-based (see section 6.3). Furthermore, we extracted significant MWEs from the immigration sub-corpora of both languages (see section 6.4), working corpusdriven again. While the keywords extracting step is going to be determinant to the following discourse comparison step (see section 7), both extracted co-occurrences and MWEs will be used as integrative elements to the interpretation and analysis of the results (see section 8).

## 6.1 Preprocessing

Before proceeding with the keywords extraction we performed an important preprocessing step: POS tagging. This was done in order to achieve firstly a lemmatisation (see section 3.2) and secondly a POS tagging. Since the goal of this step consists in the identification of the keywords related to the immigration discourse, working with lemmas instead of inflected words allowed us to reduce the amount of different word forms that were all re-conducted to their main concept, thus leading to less sparse data. The POS tagging was instead necessary for the co-occurrences extraction step (see section 6.3), since we decided to extract the co-occurrences of keywords with determinate grammatical categories.

For pre-processing we used the TreeTagger<sup>1</sup>, tool for annotating words with POS tags and lemma, developed by Helmut Schmid (Institute for Computational

<sup>&</sup>lt;sup>1</sup>http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/.

Linguistics of the University of Stuttgart). This tool, available in addition to German and Italian for other 18 languages, makes "use of a decision tree to get more reliable estimates for contextual parameters." (Schmid [1995]:2).

The TreeTagger accepts as input a plain text, as well as a tokenised one (one token per line). As shown in example (6.1), the tool returns as output a line for each token composed of token, tag, and lemma. If the lemma is not known by the TreeTagger system the general value "unknown" is assigned.

(6.1)	The	DT	the
	TreeTagger	NP	TreeTagger
	is	VBZ	be
	easy	JJ	easy
	to	ТО	to
	use	VB	use
		SENT	

In case of plain text as input, the tokenisation step is performed directly by the TreeTagger. After tagging both corpora and sub-corpora with TreeTagger, we noticed a high "unknown" rates in the Italian output. Many usual words that should be correctly recognized without any problems were not recognized. This was due to the bad tokenisation quality of the TreeTagger for the Italian language.

(6.2) dell'anno	NN	unknown
(6.3) l'Italia	NN	unknown
(6.4) d'ordine	NN	unknown
(6.5) prenderlo	NN	unknown

As shown in the examples (6.2) - (6.5), most of the errors were due to problems with apostrophes - very common in Italian in contrast to German - or clitics. To solve this problem and not compromise the lemmatisation process, we decided to tokenise the Italian text with another tagging tool for analysing language: TextPro<sup>2</sup>, "a suite of modular tools for analysis of Italian and English texts, developed at FBK<sup>3</sup>" (Pianta et al. [2008]). After tokenising the corpus with TextPro, it was successfully processed again from the TreeTagger<sup>4</sup>.

<sup>&</sup>lt;sup>2</sup>http://hlt-services2.fbk.eu/textpro/.

<sup>&</sup>lt;sup>3</sup>Fondazione Bruno Kessler, Trento

<sup>&</sup>lt;sup>4</sup>TexPro also performed the tagging and lemmatisation step. However, it does not support operations for German and uses a different Tag-set. In order to remain consistent we therefore decided to use the TreeTagger on both corpora.

From the TreeTagger output we reconstructed different files: one with just lemmas (that we used for the keywords extraction step) and one with POS tags and lemmas (used for the co-occurrences extraction) instead of the original word form. We substitute lemmas with the original inflected word forms in case of unknown outputs. We also had to deal with ambiguous lemmas of word forms that could possibly be re-conducted to more than one lemma.

(6.6) essere|stare VV:FIN

(6.7) Recht|Rechte NN

In example 6.6 the word *stato* of the sentence "in una società dove il valore della democrazia è stato usurpato dal populismo" could be re-conducted to both lemmas, *essere* or *stare*. In the same way, the word *Rechte* in example 6.7 of the sentence ", dass auch die Rechte dieses Thema endlich anerkennen" could be considered as *Recht* or *Rechte*. We disambiguated these forms according to the most frequent occurring lemma in the corpus.

## 6.2 Extracting Keywords

In order to extract the significant keywords for the immigration discourse of both countries we created the two contingency tables (with observed and expected values) and then applied the Chi-Squared test (see section 3.2.1).

For both languages we used the immigration sub-corpus as Corpus1 and the complete political corpus as Corpus2 of the contingency table (see Table 1). In this way we were able to statistically identify and extract the lemmas that are particularly relevant to the immigration discourse contrastively to the general political discourse. This method filters all the words that are common in both corpora, retrieving only the lemmas with an unexpectedly high occurrence frequency in the immigration sub-corpora.

As output we got one file per language containing lemma, Chi-Squared number and number of occurrences sorted in descending order as shown in Table 6, where *Zuwanderung* and *immigrazione* are the lemmas with the highest Chi-Squared number.

For the evaluation, following the *Table of Chi-square* (see Figure 1), we decided to select **10.82** as critical value, to obtain keywords at the 99.9% of probability<sup>4</sup>.

<sup>&</sup>lt;sup>4</sup>http://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm.

Furthermore, we considered keywords only the lemmas that occurred at least **3** times in the immigration sub-corpora. This was done to avoid possibly misspelled words or words occurring just few times in the whole corpus to be recognised as significant keywords. Keywords are reported with their Chi-squared number followed by the number of their occurrences in the political corpora.

German Output	Italian Output
Zuwanderung (1514.9568950978341, 811)	Immigrazione (3220.274000373262, 1103)
Personenfreizügigkeit (652.9121947736968, 523)	Immigrato (1859.272706815433, 637)
Asylsuchende (553.9374465656072, 304)	Clandestino (2107.0293655956903, 847)
Einwanderung (425.9267143992123, 226)	Migrare (1204.7980939261174, 416)
Masseneinwanderung (424.0418202311131, 225)	Profugo (1059.359768429323, 363)
Flüchtling (415.2253727474484, 277)	Accoglienza (1008.1222153477535, 413)
Person (400.43062735674096, 654)	Alfano (922.3385588132519, 633)
Asylgesetz (363.5860294265295, 209)	' (750.0235847612842, 2592)
Migrant (343.268515927051, 193)	Nostrum (716.8958941238261, 276)
EU (315.49200854859896, 714)	Mare (611.3736800237322, 600)
Asylbewerber (311.7001052698655, 170)	Invasione (533.8738511969416, 228)
Asylgesuch (303.83038739282165, 164)	Rifugiato (456.4182510724429, 158)
Migration (301.7952857852697, 162)	Asilo (376.05214454629794, 187)
Asylwesen (295.8744952630758, 157)	Sbarco (359.42037150144375, 138)
Masseneinwanderungsinitiative (268.3728484082283, 147)	Operazione (301.3410659115879, 269)
Arbeitskraft (248.82300392446214, 173)	Lega (299.4569499809098, 963)
Asylpolitik (245.25936585962933, 132)	Costa (290.0769549482535, 162)

#### Table 6: German and Italian most Significant Keywords

The final evaluated keywords-lists in German and Italian consisted respectively of 976 and 570 lemmas. We also calculated two other keywords-lists including the POS tags: these consisted of 1003 and 606 lemmas<sup>5</sup>. The German amount of significant words is relevantly higher than the Italian one. This is due to the high presence of compound words typical of the German language. This becomes clear just by taking a look at the first 17 keywords of both lists of Table 6. In the German list we found 6 lemmas in the semantic field of Asylum such as Asylsuchende, Asylgesezt, Asylbewerber, Asylgesuch, Asylwesen, and Asylpolitik. In the Italian list, on the contrary, we just find the lemma asilo. This because the other German words would appear as more than one word expressions in formulations like istituto d'asilo or settore dell'asilo. Furthermore, the German corpus and sub-corpus contained some articles with a minority of French sentences. Few French words such as et or de appear therefore among the German keywords<sup>6</sup>.

<sup>&</sup>lt;sup>5</sup>The difference between the two lists is due to the fact that some words were tagged differently in determinate parts of the corpora. By not considering the POS tags identical entries with different word categories would be listed as one entry.

<sup>&</sup>lt;sup>6</sup>French is an official language in Switzerland, for this reason some German articles contained few

This step is obviously very important and serves as a basis for the whole experiment. In fact, the extracted keywords are the elements that will be directly used to perform the semi-automatic comparison of the two discourses (see chapter 7).

## 6.3 Extracting Co-ocurrences

As next step, we decided to compute co-occurrences of the keywords. Starting from the extracted lists of keywords we calculated, through contingency table and loglikelihood ratio (see section 3.2.1), the words with which they significantly occur in the immigration sub-corpora.

We did not limit ourselves to understand co-occurrences strictly as two adjacent words, but we considered instead a context window of 12 (6 lemmas before and 6 lemmas after the investigated lemma). This has been done in order to find semantic and conceptual associations more than strict linguistic patterns. The first cell of the contingency table, which corresponds to the joint frequency of u and v (see Table 2) was therefore not computed for every word in the corpus, but rather for the keyword-lemmas and the lemmas in their context window. After the Loglikelihood calculation, we used the Chi-Squared table to evaluate the results, setting the critical value to **10.82**.

In this way we integrated the corpus driven-approach, accomplished in the keyword extraction step, with a corpus-based approach, since in this case we started from a pre-existing list that was automatically computed in a previous step.

We firstly decided to work just with nouns, to understand with which other nouns in the corpus they significantly occur. In order to do this, we used POS tags, thus extracting from both lists of keywords and immigration sub-corpus only lemmas with the tag NN/NE for German and NOM/NPR for Italian. Afterwards, we repeated the same operation by adding adjectives and adverbs including the tags ADJ and ADV for both languages. Finally, we moved our focus on verbs, including the tags VV/VA/VM for German and VER for Italian.

sentences or quotations in French. We decided to keep these articles in the corpus relaying on the balance of the dataset: multilingual articles should be present in both the general political corpus and the immigration sub-corpus.

As output we generated three files per language with grouped co-occurrences for each keyword:

- One containing nouns co-occurrences with others nouns (722 keywords for German and 424 for Italian);
- One containing nouns, adjectives and adverbs co-occurrences with others nouns, adjectives and adverbs (842 keywords for German and 490 for Italian);
- One containing nouns and verbs co-occurrences with others nouns and verbs (902 keywords for German and 474 for Italian);

As shown in Table 7, each output file contains the original keyword followed by its co-occurrences and their log-likelihood ratio. In the table we reported just the first 20 co-occurrences of the German keywords *Grenze* and *Zustrom*, as well as those of the Italian *mediterrano* and *confine*.

German Output	Italian Autnut
deman output	
<i>Grenze</i> : Grenzwachtkorps, 32.3097757958; Grenzübertritt, 28.3117473722; Grenzkontrolle, 23.5504611097; Kontrolle, 22.5892862538; Durchgreifen, 20.6029839152; Rhein- taler, 20.5294813802; Polizeipräsenz, 20.5294813802; Schengen-Raum, 20.2958770338; Überschreiten, 19.303051562; Wirtschaftsmigrant, 19.2361036958; Einsatzkraft, 18.2952707913; Ausgleichsmaßnahme, 18.2952707913; Assistenzdienstes, 18.2952707913; Tragbare, 17.4395452393; Militär, 17.3912930768; Bundesrat, 17.1060873754; Jahr, 16.9519758271; CVP, 16.823083654; Asyl-Moratorium, 16.7627818594; Personenkontrolle, 16.6960319886;	Mediterraneo: cimitero, 38.6331368517; sponda, 36.7167039625; Mar, 33.0532237878; governo, 32.600002612; lega, 30.2152160271; morto, 28.5308278333; nord, 26.2682539634; immigrazione, 22.9200851773; annegato, 19.4661912783; Palermo, 19.4483017693; ecatombe, 17.6714502405; trage- dia, 16.0457087038; bambino, 15.6986322422; ral- lentamento, 15.6111837965; Sami, 15.6111837965; liquame, 15.5037731389; ministro, 14.6128065395; Helsinki, 14.1433493653; pace, 14.0508575113; toccata, 13.7937580035; Annet, 13.7937580035;
Zustrom: Nordafrika, 28.9736375082; Schengen- Papieren, 27.4773242278; Schönwetter-Projekt, 27.4773242278; Ausländerbehörde, 26.1870035818; Möchtegern-Asylanten, 19.6247978928; Verfahren- szentren, 19.6247978928; Militärdienstverweigerern, 19.6247978928; Grenzwachtbehörden, 19.6247978928; 1'500, 19.6247978928; Kontrollposten, 19.6247978928; EVZ, 19.00502716; Erden, 19.00502716; Gesicht- spunkt, 19.00502716; Prioritätensetzung, 19.00502716; Rohstoffverbrauch, 19.00502716; Wirtschaftsmigrant, 18.7094040969; Asylschlendrians, 18.4694952415; Ver- antwortungsträger, 17.9979757154; Verfahrensproblem, 17.9979757154; Fachausschuss, 17.9979757154;	frontiera: controllo, 44.8562351731; lega, 30.4325794526; nord, 30.3592404239; governo, 30.2272816445; sorveg- lianza, 21.366779048; litorialità, 20.6683590142; litorale, 19.4415431018; Virilio, 19.4415431018; immigrazione, 18.7409656388; blindatura, 18.4333762522; Paul, 18.4333762522; ROMA, 17.8026167327; respingi- mento, 16.5286897398; Italia, 16.2581420662; Medici, 16.1758255931; auge, 16.1758255931; -, 15.9368116725; zattera, 15.5868716515; Baviera, 15.4872359145; teorico, 15.4872359145;

Table 7: Co-occurrences for German and Italian nouns with nouns

## 6.4 Extracting Multi-Words-Expressions

In this investigation we are not primarily interested in the idiosyncratic characteristic of MWEs, but more in n-grams that statistically significantly recur together in articles regarding the immigration topic. The frequent recurrences of n-grams in a specific context indicates a process of linguistic crystallisation. Crystallised expressions acquire specific meanings and reveals themselves perfectly suited to influence representations at a cognitive level.

To extract MWEs we proceeded again corpus driven, with a similar approach as the one used for extracting keywords (see section 6.2). We firstly calculated ngrams and then filled both contingency tables, with observed and expected values. Afterwards, we applied the Chi-Squared algorithm and evaluated the output with the Chi-Squared test setting, even in this case, the critical value to **10.82** and the minimal recurrence frequency to **3**.

We decided to compute significant 3-grams, 4-grams and 5-grams for both languages. As output we created for each language three files:

- One containing significant 3-grams with chi-squared value and occurrence frequency (782 for German and 1169 for Italian);
- One containing significant 4-grams with chi-squared value and occurrence frequency (419 for German and 581 for Italian);
- One containing significant 5-grams with chi-squared value and occurrence frequency (222 for German and 366 for Italian);

In Table 8 we reported some examples of significant 4-grams for both languages. The problem of this approach is that punctuation marks do not get "filtered out", like by the keywords extraction, causing therefore false or at least not interesting data. In fact, since article titles are often reported between quotations, part of the title containing immigration lexicon and quotation does result as significant n-gram. For this reason, both 4-gram lists of Table 8 start with 4-grams such as *gegen Masseneinwanderung*" " and, " *Immigrazione :* .

German Output	Italian Output
[gegen Masseneinwanderung " "]: 129.878740396, 69	[, " Immigrazione : ]: 644.791287405, 220
[9 . Februar 2014 ]: 117.299477692, 71	[reato di immigrazione clandestina ]: 131.875202066, 45
[Volksinitiative gegen Masseneinwanderung "]: 84.7022710138, 45	[Immigrazione : Salvini , ]: 117.222048226, 40
[Personenfreizügigkeit mit der EU]: 76.7250877052, 46	[il reato di immigrazione ]: 70.3325500133, 24
[; CVP Schweiz 2015 ]: 71.620803083, 155	[del reato di immigrazione ]: 67.4019864318, 23
[zur Steuerung der Zuwanderung ]: 56.4676576338, 30	[segretario federale della Lega ]: 59.5549836668, 73
[und Begrenzung der Zuwanderung ]: 52.8561821281, 29	[federale della Lega Nord ]: 58.1179066492, 71
[an Leib und Leben ]: 51.772170827, 31	[il deputato della Lega ]: 37.4029293945, 80
[Steuerung und Begrenzung der ]: 48.9385157365, 26	[in materia di immigrazione ]: 32.2354992659, 11
[machen uns stark für ]: 45.0512221806, 39	[il permesso di soggiorno ]: 29.3049816528, 10
[Wir machen uns stark ]: 42.6619943792, 37	[. Siamo la casa ]: 23.4439570346, 8
[harte , aber faire ]: 42.099604823, 25	[Siamo la casa di ]: 23.4439570346, 8
[Zuwanderung in die Schweiz ]: 39.690463647, 22	[la casa di chi ]: 23.4439570346, 8
[in die Schweiz kommen ]: 37.8102001051, 21	[vera e propria invasione ]: 21.2530657272, 8
[der Personenfreizügigkeit mit der ]: 36.1794400195, 21	[gestione dei flussi migratori ]: 21.2530657272, 8
[die Personenfreizügigkeit mit der ]: 34.3055760683, 20	[bloccare le partenze , ]: 20.5134500295, 7
[Leib und Leben bedroht ]: 34.0502973231, 19	[vitto , alloggio , ]: 20.5134500295, 7
[eine harte , aber ]: 32.1707248211, 18	[per bloccare le partenze ]: 20.5134500295, 7

Table 8: Significant 4-Grams for German and Italian

Despite this fact, significant n-grams are very useful for our analysis. In contrast to co-occurrences they serve exactly to analyse linguistic patterns. These often hide defined and well recognizable cognitive patterns such as the one discussed in chapter 2.

# 7 Comparing Discourses

In this section we describe the discourse comparison process. This is primarily based on a direct comparison between the two keywords-lists extracted in the previous chapter (see chapter 6.2). To perform the comparison we trained a SMT system which generated, among other outputs, files containing translation pairs stored with their translation probabilities.

The process consists mainly in training the TM of our SMT system with a big German-Italian parallel corpus. As parallel corpus we used data from *Europarl* and *bulletin4corpus* as explained in section 5.2. The complete corpus consisted of 1'909'069 lines (1'832'052 of *Europarl* and 77'017 of *bulletin4corpus*).

The best way to possibly cover the immigration topic was to create a more or less domain-specific parallel corpus. Therefore, we decided to use both corpora, under the assumption that the sum of these would mostly cover both political discourses about immigration. Concretely, we aimed to match with our translation system as many keywords as possible. Therefore, even country-specific lexicons such as proper names of politicians, parties, laws, or national-specific expressions were important. *Europarl* thematically cover the immigration discourse, containing the transcription of sessions of the European Parliament. Furthermore, since Italy is part of the EU, it possibly contains also Italy-related issues and consequently lexicon. *bulletin4corpus* does include the immigration topic as well, since many articles are about society. Moreover, it is completely Switzerland related: this is relevant since Switzerland, as not EU-member, is not directly represented in *Europarl*. Of course, since Germany and Austria are part of the EU, *Europarl* contains the German standard immigration lexicon. However, the addition of a Swiss-specific corpus entails the coverage of Swiss-specific vocabulary.

The training of the TM was preceded by a preprocessing step and followed by the bi-directional comparison between the two keywords-lists.

## 7.1 Preprocessing

In order to possibly match any keyword of the extracted lists we had to obtain a parallel corpus, which contained sentences made of lemmas instead of word forms.

For the *Europarl* part of the corpus, we proceeded exactly like for our political corpora and sub-corpora. We tagged both corpora files with Treetagger<sup>1</sup> and recreated, after the disambiguation of ambiguous lemmas, two sentences aligned texts with lemmas instead of word forms. The *bulletin4corpus*, on the other hand, already provides POS tags and lemmas (see section 5.2.2). Therefore, we extracted exclusively lemmas from both XML files and generated two sentence-aligned files using the bulletin4corpus-alignment file.

As last step, we merged both corpora in order to have one big file per language containing all the data.

## 7.2 Translation Model

To train the TM we used the tools provided by  $Moses^2$ . Moses is an open source implementation for SMT, developed by Hieu Hoang and Philipp Koehn and available since 2005. It allows training SMT systems with external parallel corpora. Moses has two main components: "the training pipeline and the decoder" (Koehn [2016]:12). The training pipeline consists in a collection of tools that are mainly written in Perl and C++, and creates an SMT system that takes plain parallel text as input. The decoder consist of a single application written in C++ that "given a trained machine translation model and a source sentences, will translate the source sentence into the target language" (Koehn [2016]:12).

The training pipeline provides some scripts in order to perform three preprocessing steps that improve the SMT results: the first step consists in cleaning the corpora; the second in filtering out all the sentences longer than 35 characters; and the third in lowercasing the corpora. After these three operations the corpus consisted of 1'216'043 lines.

The training script performs the word alignment and creates the lexical translation files as well as the phrase table. The word alignment is done with  $\text{GIZA}++^3$ 

<sup>&</sup>lt;sup>1</sup>The Italian corpus was also in this case tokenised with TextPro.

<sup>&</sup>lt;sup>2</sup>Cfr. http://www.statmt.org/moses/

<sup>&</sup>lt;sup>3</sup>http://www.statmt.org/moses/giza/GIZA++.htm.

which is based on the IBM models (see chapter 4).

We performed the training in both directions, initially setting German as first and Italian as second language parameter, and then inverting them. The output results were slightly different depending on the translation direction. The most interesting output for us is the one exemplified in Table 9, that contains possible lexical pairs with their translation probabilities. Table 9 shows the possible translation of *Zustrom* and *afflusso*. We reported exclusively a selection of translation examples, including the highest probable pairs.

German to Italian	Italian to German
zustrom inondare 0.0135135	afflusso zufluß 0.2727273
zustrom ondata 0.0099010	afflusso migrantionsströmen 1.0000000
zustrom ingresso 0.0037879	afflusso gesamtkapitalzufluss 0.5000000
zustrom arrivo 0.0230906	afflusso fluchtbewegung 0.2000000
zustrom immigratorio 0.1000000	afflusso andrang 0.0769231
zustrom invasione 0.0050761	afflusso hoch 0.0001599
zustrom afflusso 0.2645740	afflusso massenansturm 0.2000000
zustrom esodo 0.0117647	afflusso strom 0.0051282
zustrom flusso 0.0225303	afflusso zuzug 0.2941176
zustrom affluenza 0.0130719	afflusso passagierzahl 0.0370370
zustrom segnale 0.0003186	afflusso massenzustrom 0.2380952

Table 9: Lexfile Ouputs

## 7.3 Keywords Comparison

As fist step we performed an output synchronisation, merging for each language the lexical files obtained with both translation directories. At this point, we created for each lexical file a translation file, containing for each word in the source language only the 5 most probable translations in the target language. In Table 10 we report the examples from the translation file with Italian as source and German as target language from *convalescenza*, *attendibilità*, *consumatrice*, *stregone*, *intenzionalmente*, *caseificio*, and *regolamentativa*. As shown in the table, the correct translation is not always in first position. Therefore, we decide to include the first 5 results, considering that some of them are going to be repeated<sup>4</sup>.

<sup>&</sup>lt;sup>4</sup>This is due to a different translation probability number in the two lexical files that we merged.

Italian Word	German Translations
Convalescenza	Genesungsaufenthalt, Gesundheitsnetz, Nachsorge, Genesung, Sprengen
Attendibilità	Datenqualitaet, Wahrheitsgehalt, Verbuergen, Zuverlässigkeit, Berechenbar
Consumatrice	Verbraucherin, Verbraucherin, Straßentransportunternehmen, Menge, Menge
Stregone	Hexendoktoren, Zauberlehrling, Hexenkinder, Munter, Jonglieren
Intenzionalmente	Gewollter, Pornovertreiber, Manipulieren, Vogelstraußpolitik, Willentlich
Caseificio	Molkerei-, Fleischereibetriebe, Milchfabrik, Milchviehbetrieb, Ungeliebt
Regolamentativa	Regulierungspolitik, Ordnend, Vorschrift, Vorschrift

### Table 10: Translation File Italian to German

Afterwards, we had to check if the translation files included the keywords of both lists<sup>5</sup>. This step is necessary because the lemmas not included in the translation files cannot be translated and used for the confrontation. To do this we used the keywords-lists calculated without PoS tag (see section 6.2).

Out of 976 German words:

- 783 words were matched and saved in a file with their translation;
- 193 words were not matched and saved in a separate file;

Among the German keywords that were not matched appear mostly words that are specifically related to the Swiss discourse such us Ventilklausel, Sommaruga, Inländervorrang, SVP, or CVP. These are not highly relevant since they are for sure not present in the Italian discourse. Furthermore, also words that are still conjugated such us Bilateralen, Freizügigkeitsbkommens, Asylbereichs, or Asylsuchenden figure as unmatched words. These words are matched however in their nominative, masculine, singular form. Important terms that were missing from the German matched list are:

- Asylchaos • Zuwanderungsbegrenzung
- Ausländerkriminalität
- Sozialmiqrant
- Asylindustrie
  - Familiennachzug

These are all compound words. As we already mentioned, German disposes of a large amount of compounds and it even provides the possibility to create neologisms by creating new compounds. Compounds are created by joining different already existing words together. This leads to an increase of the vocabulary size and to sparse data problems (Popović et al. [2006]). It is, in fact, very difficult to cover all

<sup>&</sup>lt;sup>5</sup>In order to match the translation file the keywords-lists were lowercased.

these cases with any parallel corpora, and to built a system capable of translating these terms.

Out of 570 Italian words:

- 505 words were matched and saved in a file with their translation;
- 65 words were not matched and saved in a separate file;

Even in the case of Italian, some specific lexicon related to the country consisting of words such as *Triton*, *Renzi*, *carroccio* (surname given to the Lega Nord party), *leghista*, *Italicum*, or *Viminale* were not matched. Regarding relevant missing terms we only identified *scabbia* among the terms that were not matched.

At this point we cross-matched the two lists with matched words in the respective source language and the five translations in the target language. The output consisted of one keyword in source language per line with its translation in target language, followed by its five translation possibilities. Keeping the translation possibilities allowed us to follow why two words were matched as translations.

Italian Word	German Translations + Stored Translations
Emergenza	Nothilfe ['conforto', 'emergenza', 'emergenza', 'soccorso']
Vita	Leben ['andere', 'vita', 'vita', 'vitavite']
Nostro	Unser ['ghisa', 'diventado', 'vun', 'nostro']
Immigrazione	Zuwanderung ['dell02bcaccesso', 'immigrazione', 'clandestini', 'migrazione']
Agricoltura	Landwirtschaft ['agropastorale', 'agricolto', 'agricoltura', 'agricoltura']
Espellere	Ausweisen ['rcd-goma', u'info', u'5-510', espellere']
Chiesa	Kirche ['chiesa', u'chiesa', u'metodista', santificare']

### Table 11: Translation Match Italian to German

For the matching direction Italian to German we did not encounter any problems. Out of the 505 Italian keywords:

- 128 words were matched and stored in a file with their translation;
- 377 words were not matched and stored in a separate file;

For the opposite matching direction the process became more complex. Many translation possibilities of the Italian matched words were compounds. To provide some examples: the word *popolazione* is "translated" with *Eubevölkerung, Bevölkerungszuwanderung, papuanische,* and *Bienenpopulationen*; the word *paura* with *Grundängste, Angstbewältigung, Globalisierungsblues,* and *nanogrösse*; and *protezione* with *Umweltschutzpriorität, einführen, herabziehen, Schutzgewährung,* and *Schutzauftrag.* Most of these words contains the correct translation, but are

too specific since they are compounds. For this reason, they were not matched within our system. The presence of these too specific words is due to the fact that generic words like *Schutz* or *Bevölkerung* mostly recur in the translation corpora in sentences regarding specific topics. In these cases, where Italian uses an expression with preposition like "protection against something", German directly uses compounds. To solve this problem, we used a tool for splitting German compounds: jWordSplitter<sup>6</sup>. jWordSplitter is a Java library that splits German compounds into their smaller components. The word *Erhebungsfehler* for example is split in *Erhebung* and *Fehler*. This tool works best with nouns, but it also works with adjectives and verbs.

As first step, we used the initial Italian keywords file (containing keywords followed by their not split translation possibilities) and matched **51** words out of the 783 initial German keywords. Afterwards, we used jWordSplitter on the German translation possibilities of the Italian file, and repeated the process again, matching other **87** keywords. Totally:

- 137 words were matched and stored in a file with their translation;
- 646 words were not matched and stored in a separate file;

The intersection between the two files containing matches in both direction consists of **176** keywords. These are shared between both the Swiss-German and the Italian political discourse. In Table 12 we reported some examples of the output containing intersecting keywords, all stored in the translation direction German to Italian and in alphabetical order.

German	Italian
Europa	Europa
Eu	Ue
Februar	Febbraio
Fliehen	Fuggire
Fliehen	Scappare
Flucht	Fuga
Flüchtling	Profugo
Forschung	Ricerca
Freizügigkeitsabkommen	Circolazione
Fremdenfeindlich	Xenofobo
Gegen	Contro

Table 12: Intersecting Keywords

<sup>&</sup>lt;sup>6</sup>http://www.danielnaber.de/jwordsplitter/index\_en.html

The totals of unmatched keywords for German and Italian were 841 and  $442^7$ .

 $<sup>^{7}</sup>$ Totals contain keywords that were not matched in both the lexical and the translation files.

# 8 Analysis and Interpretation

In this section we will perform the analysis and interpretation of the previously extracted data. The focus will be on the results of the keywords comparison (see chapter 7), which will be when possible integrated with the extracted co-occurrences and MWEs<sup>1</sup> (see sections 6.3 and 6.4). In order to have an analysis as complete and consistent as possible, we will always report one or more recurrence examples of the investigated term in its context. Along with them we are going to include party and date of the press release where the term appears. Although the analysis is not going to be party-dominant - meaning that we do not intend to highlight the different argumentative strategies actuated by the different parties - we decided to give note of some party-related issues. This is necessary, because, as we will see, the argumentative topoi of immigration discourse are mostly developed in right-wing-parties speeches.

We are going to use the term *migrant* with the generic meaning of a person who leaves his country or region to settle in another one (Glo [2012b]) for addressing both immigrants and asylum seekers. The latter are going to be referenced as asylum seekers or refugees only when we will exclusively talk about them. For interpreting the data we are going to consider argumentative strategies such as legitimation and coercion through the analysis categories previously introduced (see section 2.2).

<sup>&</sup>lt;sup>1</sup>MWEs are always going to be reported between square brackets.

## 8.1 Thematic Comparison

To have an idea of the differences existing between the Swiss-German and the Italian discourses we started by identifying different topics. In order to identify the topics we performed a superficial semantic comparison of the recurring and not recurring keywords. In this context we also took into consideration MWEs about the different topics.

German	Italian	MWEs
Ausländer	Straniero	[Ausländern in der Schweiz]
Integration	Integrazione	
Zuwanderung	Immigrazione	[Steuerung und Begrenzung der Zuwanderung] [einen generellen Stopp der Zuwanderung], [il reato di immigrazione clandestina] [il problema im- migrazione]
Fremdfeindlich	Xenofobo	
Einwandern	Immigrare	



Table 13 gives an overview of the general terms present in both discourses. Both lists contain the generic concepts that we would expect in discourses about immigration. In addition to these widespread terms, we recognized two distinct sub-discourses as shown in Tables 14, and 15:

- European migrant crisis (asylum seekers);
- European intern immigration;

Both nations share the topic of European migrant crisis. It is, however, much more developed in the Italian discourse, as confirmed by both the higher quantity of related keywords and the frequent presence of MWEs. The Swiss-German discourse focuses on the issues of refugees, repatriations, and people-smugglers. However, it does not include as much as the Italian one the topic of landing by sea, which in turn implies a series of other issues such as drowning of migrants, first aid, and tragedies at sea.

German	Italian	MWEs
Aufnahme	Accoglienza	[sistema di accoglienza]
Asyl	Asilo	[in der Asyl- und Ausländerpolitik] , [il diritto di asilo]
Rückführung	Rimpatrio	
Illegale	Clandestini	[illegale Einwanderer ,] , [centinaia di migliaia di clandestini]
Mittelmeer	Mediterraneo	
Schlepper	Scafista	
Rifugiato	Flüchtling	[Flüchtlinge aus Syrien], [status di rifugiato]
Grenze	Confine, Frontiera	[Grenzen wieder systematisch zu kontrollieren], [difesa dei confini] [la chiusura delle frontiere]
Küste	Costa	[le nostre coste]
Terrorist	Terrorista	
	Barcone	
	Nostrum, Triton	[Mare Nostrum e Triton]
	Mare	[morti in mare]
	Naufragio	
	Strage	[strage di migranti]
	Schengen	
	Bilaterali	
	Circolazione	
	UE	

Table 14: European Migrant Crisis

Instead, the topic of European immigration is exclusively part of the Swiss-German discourse, as shown in Table 15. This is due to the large number of European migrants moving to Switzerland for financial and work-related reasons. The terms *Schengen, bilaterali, circolazione (libera)* and *UE* of the Italian keywords-list refer to the common management at European level of the European migrant crisis, and not to the EU internal immigration. For this reason, they were allocated in the previous table even if some of these concepts are the literal translation of ones recurring in the Swiss-German discourse.

German	Italian	MWEs
Schengen-Abkommen		
Bilaterale		[bilateralen Verträge mit der EU ]
Personenfreizügigkeit		[der Personenfreizügigkeit mit der EU] [negativen Folgen der Personenfreizügigkeit]
EU		[Beziehungen zur EU ] [Einbindung in die EU ]
Eu-ausländer		
Fachkräftemangel/Fach- frakf		[Fachkräfte aus dem Ausland ]
Kontingent		[jährliche Höchstzahlen und Kontingente]
Lohndumping		

Table 15: European Intern Migration

Summarising, while the Swiss-German discourse includes both topics (European migrant crisis and European internal migration), the Italian discourse is only about the first one. This disparity is explained by the different historical and sociopolitical situation of the two countries. Switzerland, as an economically strong country that is not part of the European Union but largely dependent on the economical and political relations with it, is faced with the issue of European immigration. Furthermore, because of its geographical position in the middle of Europe, the country is not directly confronted with the landings of refugees. Italy, on the other hand, was hit by the economic crisis, and is one of the main access to Europe by sea, second only to Greece. For this reason Italy is directly confronted with the so called European migrant crisis.

## 8.2 Legitimation

As previously discussed, legitimation is a necessary operation in order to achieve coercion (see chapter 2.2.1). To possibly influence the opinion of the text consumer, the text producer must have authority. We started the legitimation investigation by searching for terms that could indicate the actuation of the epistemic strategy (Chilton [2004]), such as *politician*. Epistemic claims are often supported by statistics and sources. Therefore, we also searched for objectivity markers like *fact*, *number*, or *dates*.

The term *Politiker* is present in the Swiss-German list, however, it is not used in the way that we would expect in order to perform the epistemic strategy.

According to the co-occurrences of Table 16 the term *Politker* is not related to the epistemic strategy. Instead, it is used to implement a reverse strategy, actuated mainly by the SVP party. This strategy involves popular willing and appeals to

Co-occurrences	Context
Machtinteresse	Das Volk stimmt meist für das Gemeinwohl. Es stellt den Menschen in den Mit- telpunkt. Wenn nur noch Politiker entscheiden, dann spielen Machtinteressen und die Beeinflussung durch Filz sowie persönliche Vorteile oder Privilegien eine wichtigere Rolle. SVP 14.04.2015
Schultern	Es sind die Frauen, Männer und Kinder dieses Landes, welche die Last des Asyl- und Flüchtlingswesens auf ihren Schulten tragen müssen. Nicht der Bundesrat, die Beamten oder die Politiker! SVP 27.08.2015
Gelehrte, Volksrechte	Richter und Gelehrte irren Linken Politikern sind die Volksrechte schon lange ein Dorn im Auge. [] Ach, wie bequem wäre es doch, wenn nicht noch der Bürger das letzte Wort hätte. [] Das Volk stimmt meist für das Gemeinwohl. SVP 14.04.2015

Table 16: Co-occurrences of *Politiker* 

popular sovereignty, traditional element of the Swiss electoral system, which is based on direct democracy. The politician figure is coupled with power interests and described as an intellectual, which is not part of the common people but of a separate elite. This strategy is very effective and produces a strong separation between the so-called "classe politique" and the SVP politician, which claims, on the contrary, to truly represent the people. In support to this argument we analysed some recurrence contexts of the term *Classe politique*, which also resulted as keyword<sup>2</sup>.

- (8.1) Damit hat die Mehrheit von Volk und Ständen der Classe politique klargemacht [...] . SVP 01.05.2014
- (8.2) Die Classe politique, der Bundesrat und die anderen Parteien schauen vor dieser Problematik weg, es wird behauptet, die Schweiz habe kein Ausländerproblem. SVP 27.07.2010

The SVP politician uses popular legitimation obtained through succeeded popular initiatives to situate itself in an irrefutable position of guarantor of popular willing, since the direct democracy is deeply-rooted in Switzerland. Consequently, SVP adversaries are framed as an anti-democratic elite, which acts against popular willing. The politician figure is not represented as a super entity that knows what is best for the population, since it is the population itself that is able to effectuate the best decisions. This strategy contributes in creating a clear dichotomy between an us, SVP-politicians and population, and a them, corrupted socialist elite. This is supported by the claims of objectivity resulting from the co-occurring of *Zahl*, found in the keywords-list with verbs such as *sprechen* or *zeigen* (examples 8.3 - 8.5).

<sup>&</sup>lt;sup>2</sup>Both *classe* and *politique* recurred in the Swiss keywords-list

- (8.3) Die Zahlen zeigen, dass selbst in einer Phase mit steigender Arbeitslosigkeit ein Heer von Arbeitskräften mit ihren Familien in die Schweiz einwandert und auch in der Krise hier bleibt. SVP 24.08.2010
- (8.4) Doch die Vox-Zahlen sprechen eine deutliche Sprache. SVP 03.04.2014
- (8.5) Die neusten statistischen Zahlen zeigen demgegenüber, wie wichtig es wäre, endlich konsequent gegen kriminelle Asylbewerber vorzugehen. SVP 26.03.2012

The objectivity of numbers is often used to prove the legitimation of a certain linguistic expression. It does not matter whether the source of the numbers is declared, an affirmation supported by numbers always seems more effective.

*Politico* does not recur in the Italian keywords-list. Neither the traditional epistemic strategy, nor a reversed one, are actuated. We then searched for objectivity markers and found only the term *dati*, which co-occurs with predicates such as *dimostrare*, *dare*, and *parlare*.

- (8.6) Da anni la Lega continua a ribadire di stare attenti perché, confuso tra i profughi, può arrivare chiunque in questo Paese, e i dati ci hanno dato ahimè - spiacevolmente ragione. Lega 03.12.15
- (8.7) I dati parlano chiaro, sono pochissimi quelli che hanno vera necessità di protezione. Lega 11.12.15
- (8.8) L'aumento degli sbarchi è infatti dovuto alla destabilizzazione dello scenario internazionale e, come dimostrano i dati, è del tutto scollegato dalle scelte di politica interna. PD, 26.08.2014

Unspecified data are used, as shown in the examples 8.6- 8.8, to support the claims made by the different parties. The Italian legitimation strategy is exclusively characterised by the usage of objectivity markers. Formulations that combine terms like *numbers* or *data* with the verb *speak* are very interesting from a linguistic perspective. The objective markers are personified through the use of verbs usually related to humans. The legitimation effect of this action is strong: it is not a matter of a subjective opinion, rather of an "objective" claim.

## 8.3 Coercion

In this section we will analyse the implementation of different coercive strategies in the discourses. Firstly, we will focus on the different ingroup - outgroup representations. Then we will concentrate on emotive and cognitive coercion.

### 8.3.1 Ingroup Vs. Outgroup

As we previously explained, the process of dividing the world into "them" and "us" through social categorization is determinant to achieve every kind of coercion. Deictic elements such as possessive nouns are fundamental for this strategy. The possessive *our* appears in both lists. After analysing the recurrence of *unser* and *nostro* among the MWE's, we could outline a table (Table 17) to show their different use.

MWE	Context
Freiheit: [unsere Freiheit ,]	Die heutige Masslosigkeit bei der Zuwanderung gefährdet unsere Freiheit, Sicher- heit, Vollbeschäftigung, unser Landschaftsbild und letztlich unseren Wohlstand in der Schweiz. SVP 25.11.2013
Regeln: [an unsere Regeln]	SVP Frauen Schweiz wollen mehr Sicherheit, indem diejenigen Ausländer, die sich nicht an unsere Regeln halten und schwere Verbrechen begehen, konsequent aus der Schweiz ausgewiesen werden können. SVP 16.11.2010
Land: [in unserem Land kommen]	Das Plakat zeigt aber auch exemplarisch auf, dass es eng wird in unserem Land. SVP 10.12.2013
Wohlstand: [unseren Wohlstand in der Schweiz]	Der Druck aus dem Ausland auf unsere Unabhängigkeit und Souveränität und damit letztlich auf unseren Wohlstand wird von Tag zu Tag grösser. SVP 14.12.2011
Coste: [sulle nostre coste]	altre 800 mila persone arriveranno a breve sulle nostre coste. Lega 24.06.2014
Frontiere: [difendere le nostre frontiere]	È tempo di pensare agli italiani e di chiudere le nostre frontiere. PDL 03.08.2015
Casa: [in casa nostra]	il terrorismo islamico è a due passi da 'casa nostra'. PDL 24.02.2015
Disoccupati: [i nostri disoccupati]	Ma prima ci sono i nostri disoccupati da aiutare Lega 13.03.2015
Gente: [la nostra gente]	Si tagli sull'immigrazione e si pensi, una buona volta, alla nostra gente. Lega 01.10.2014
Cittadini: [i nostri cittadini]	prima di tutto vengono i nostri cittadini. Lega 21.07.2015
Spese: [a nostre spese]	Ancora a nostre spese. Lega 13.05.2015
Città: [nelle nostre città]	le nostre città non sono più in grado di ricevere nessuno. Lega 03.04.2015

Table 17: Possessive Nouns: unser, nostro

Table 17 shows a difference in the usage of the pronoun. In the Italian discourse *nostro* is used in a more territorial way to define two things. On one hand it defines territorial entities such as borders, coasts, cities, and houses. On the other it also defines directly the ingroup members, for example in expressions such as *in nostri* disoccupati, *i nostri cittadini*, *la nostra gente*. The Swiss-German discourse, instead, has the tendency to define as *unser* more socio-cultural values such as freedom, laws,

and prosperity. However, in both discourses, its usage aims to signal and underline the clear delimitation between the ingroup and the outgroup. The usage contexts of the Italian discourse demarcate the difference more strongly since the outgroup members are directly implicated. This is also proved by the presence of the plural third-person pronoun *loro*, which is missing from the Swiss-German list.

Loro recurs exclusively in the MWE [a casa loro], in sentences like

- (8.9) Aiutiamoli a casa loro. Lega 14.09.2015
- (8.10) Poi possiamo aiutarli a casa loro trasformando il Mediterraneo in un crocevia commerciale e di pace. UDC 27.09.2015

The concept of *local aid* or *assistance on site* is also deeply-rooted in the Swiss-German discourse, as the MWE [Hilfe vor Ort] (Menschliche Asylpolitik heisst Hilfe vor Ort. SVP 19.03.2013) shows. In both languages the concept of local aid is based on the deresponsibilisation of Europe regarding the migrant crisis. The Italian expression is once again more incisive, since, by making the possessive pronoun explicit, it restates the ingroup - outgroup distinction.

To show once more the strong dichotomy between ingroup and outgroup we analysed metonymies used to reference migrants. This was done to show the pragmatic function of these references compared to the official reference categories of the two languages illustrated below:

- CH: Asylbewerber, Flüchtling, Asylsuchende, Kriegsflüchtling, Migrant, Zuwanderer, Einwanderer, Immigrant, Aufenthalter, Kontingentsflüchtling (Glo [2012a]);
- IT: Immigrato, Rifugiato, Richiedente (asilo), Profugo, Migrante (Glo [2012b]);

Despite the amount of official terms to reference migrants parties press releases often use other terms, actuating a clear referential sub-strategy. The active choice of using referential forms that differ from the official ones has consequences especially at the level of pragmatic.

Reference Type	German	Italian
Legal Status	Illegale, sans-papier : Immer mehr Illegale [] gelangen durch Schlepper in die Schweiz. SVP 13.10.2015, Grüne sind solidarisch mit Sans-Papiers. Grüne, 28.06.2010	<i>clandestino, irregolare</i> : Come la Lega da sem- pre sostiene, gli irregolari devono fare ritorno al loro Paese. Lega 02.07.2015, Ogni clandestino riceve più del doppio di una pensione minima. Lega 24.09.2014
Immigration Motivation	Wirtschaftsmigrant,Sozialmigrant,Wirtschaftsflüchtling:So wird dieAnziehungskraftderSchweiz für []Wirtschafts-undSozialmigrantenerhöht.SVP 25.09.2015,Denn bei diesenAsylsuchenden handelt es sich in ersterLinieum Wirtschaftsflüchtlinge.FDP 17.03.2012	
Illegitimacy	Scheinasylant, Asylant, Asylmissbraucher: Die Scheinasylanten in Neuenburg suchen nichts anderes als bessere Lebens- und Ar- beitsbedingungen. SVP, 04.05.2011, [] dass die Schweiz Asylmissbraucher nicht toleriert. SVP 25.06.2015	
Ingroup Affiliation		Extracomunitario : Triton [] illude migliaia di extracomunitari che raggiungere l' Italia sia cosa facile. PDL 12.02.15, lunghi e violenti monologhi contro extracomunitari e Rom, con raccapriccianti ironie sulle morti in mare di mi- granti. SEL 10,11,14
Emotiv Status		<i>disperato</i> : l'Italia continua a incentivare la partenza di migliaia di disperati. Lega 03.08.2015, l disperati che cercano la salvezza li dobbiamo salvare. SEL 24.06.2015

#### Table 18: Referential Metonymies

Table 18 shows a general tendency, common to both discourses, to define the migrant not as a person but through metonymical processes. Metonimical processes are always related to a particular semantic field. We identified metonymies based on the legal status, on the immigration motivation, on the legitimacy of immigration, on ingroup affiliation, and on the emotive status of the migrant. In particular, in Italian we noticed the change of category of many adjectives, which turn into nouns maintaining their originally attributive meaning. The first and most evident pragmatic effect is the creation of a permanent bonding between the individual and its condition of illegality (*clandestino, irregolare*), of exclusion (*extracomunitario*), and of negativity (*disperato*). The dehumanization of the migrants - defined only through negative attributes - contributes to the creation of prejudicial inferences in respect to the subjects. These negative attributes are not considered as characteristic of single people, but become specific of the migrant category as a whole, combining numerous humans with individual history, situations, and conditions and erroneously defining them.

The Swiss-German discourse shares the legal metonymy (*Irregulare*) present also in the Italian discourse. However, its focus is more on references based on the illegitimacy status (*Scheinasylant, Asylmissbraucher*), and on the motivation of immigration (Wirtschaft- und Sozialmigrant). Both contribute in establishing the idea of migrants that seek asylum for dubious reasons. The explicit expressed immigration motivations in the compound are not essential, and always negative for the ingroup. Asylum seekers are not described as in need, but as people only wanting to improve their economical situation at expenses of ingroup members. Also the term Asylant, which may seem neutral, hides a strong negative meaning. Wengeler<sup>3</sup> (Wengeler [2003a]) defines the term as a defaming designation. Firstly, the ending "ant" often appears in German in disparaging nouns such as Simulant, Spekulant, Querulant, or Bummelant. In addition, this ending is often used in compounds with a clearly negative connotation, such as *Scheinasylant*, or *Wirtschaftsasylant*. Consequently, the ending has undergone a process of stigmatisation. In fact, the European Asylum and Migration Glossary ([Glo, 2012a]) does not include Asylant among the official terms related to immigration. The use of all these terms already introduces the topos of exploitation that, as we will see in what follows (see section 8.3.3), is one of the most prominent characteristic of Swiss-German discourse. Migrants are presented as bogus and cheater that only aspire to improve their socio-economic conditions at the expense of the ingroup.

The fact that even left-wing parties such as *Grüne* and *SEL* use some of the analysed referential metonymies shows that these terms are not necessarily perceived as negative by the text producer. However, even if the referential metonymy is used in a context that differs from the right-wing-parties one, the impact in the text consumer is the same. In fact, the use of these terms is always related to a negative perception. Furthermore, over the time, terms used to single occurrences become the ontological propriety of the whole migrants category.

The last analysis category that we want to investigate regarding the ingroup - outgroup topic is the positive and negative face theory. When comparing the recurring terms contained in both keywords-lists, and their co-occurrences in the corpora, we noticed a different approach in the positive and negative face addressing strategy. While the Swiss-German discourse is characterised by a positive face addressing strategy - understood in Chilton's way (see section 2.2.2.1)- the Italian discourse is marked by the mistreating of the negative face.

To investigate the positive face addressing, we started by searching for *Switzerland* and *Italy* in the respective keywords-lists. This was done under the assumption that the country name might be the most direct term, beside the already investigated possessive noun, to appeal to the ingroup. Regarding the Swiss discourse, a

<sup>&</sup>lt;sup>3</sup>The linguist dedicates considerable space to history and reception of this specific term in the German public discourse between 1960 - 1985.

positive face addressing strategy is easily identifiable. The term *Schweiz* is in the list and co-occurs with *Erfolgsmodell*, *Zielland*, *Sprache*, *Wohlstand*, and *Attraktivität*, also present in the list.

Co-occurrence	Context	
Schweiz - Erfolgsmodell	Die Schweiz ist ein Erfolgsmodell. CVP 30.10.2013, Die Schweiz hat Zuwanderung weil sie ein Erfolgsmodell ist. CVP 01.08.2015	
Schweiz - Zielland	Die Attraktivität der Schweiz als Zielland für Asylsuchende muss gesenkt () werden. SVP 14.01.2015	
Schweiz - Sprache	Zu Recht, denn die Schweiz hat sich schon immer durch verschiedene Sprachen, Kulturen und Religionen ausgezeichnet. SP 30.07.2015	
Schweiz - Wohlstand	Zudem haben die Verbandsvertreter offenbar vergessen, welche Qualitäten der Schweiz die Basis unseres Wohlstandes ausmachen. SVP 06.01.2014	
Schweiz - Attraktivität	Zudem ist die Attraktivität der Schweiz als Asylland deutlich zu senken. SVP 04.11.2014	

Table 19: Swiss Positive Face Addressing

As we can see from the examples in Table 19, Switzerland is presented as a successful model that currently detain a high level of prosperity and wealth. Even if the country seems to be put in danger by various issues - which can be immigration or suspension of bilateral agreement on free movement depending on the party that formulates the concern - the positive face of the text consumer is brought into play by mentioning the current condition of the country and of its inhabitants. Defining Switzerland as target or destination country contributes again to the positive face addressing strategy, entailing its attractiveness. The country is presented as one with a "too good" assistance service.

Italia does not appears in the Italian list, as well as *prosperità* or *attrattività*. However, we found terms such as *paese* and *cittadino* that can be useful for our purpose. Among the investigation of these words and their co-occurrences there is no trace of positive face addressing, which leaves room to a strategy that mistreat the negative face of the text consumers.

Co-occurrence	Context
Cittadino - tasche	I 2,2 miliardi spillati dalle tasche dei cittadini per dare vitto, alloggio, sigarette e cellulari ai clandestini pretendiamo che vadano alla nostra gente, ai nostri cassintegrati, ai nostri disoccupati, alla sicurezza, in un Paese diventato preda di bande criminali che assaltano case al ritmo di una ogni due minuti. Lega 28.04.2015
paese - vittima	Il fenomeno immigratorio è sfuggito a ogni controllo e il paese è vittima di un'autentica invasione. Lega 19.06.2014
Cittadino - tagli	La risposta del governo Renzi e dell'SVP alla richiesta di maggiore sicurezza dei citta- dini sono gli svuotacarceri e i vergognosi tagli alla sicurezza che stanno perpetrando a danno dei cittadini onesti e delle forze dell'ordine. Lega 01.04.2014



As Table 20 shows, the situation in the Italian discourse seems to be the opposite of what Chilton (Chilton [2004]) describes as negative face addressing. The attempt to minimise the problems does not persist. On the contrary, we notice the negative face mistreating, which is carried out by underlining the current critical social and economical situation in Italy, and by putting the blame on immigration. The personal freedom of the ingroup members is under threat : they are described as blameless victims of the government that introduces continuously new cuts.

After analysing the usage of possessive pronouns, referential metonymies, face addressing and mistreating theories, in regard to the ingroup - outgroup delimitation, we can already see some trends in the discourses. While the Italian discourse creates the differentiation using concrete elements (like possessions or territory), and blames immigration for the problematic socio-economical situation of the country, Switzerland is fairly more self-concentrated. The outgroup members are characterised and defined only in relation to the ingroup (they are presented as ingroup exploiters), whose positive face is addressed by describing the country as a successful model.

### 8.3.2 Emotive Coercion

To investigate the use of emotive coercion we started searching for terms in our lists that could certify the presence of the metaphors that Saiz de Lobado and Bonomi (Saiz de Lobado and Bonomi [2012]) described as typical for the immigration discourse (see section 2.2.2.2). In Table 21 we grouped our findings. We recognised the metaphorical fields of natural disaster, act of war and biblical references. These all contribute to dehumanise migrants, painting them as a threat to the host country.

The most popular metaphorical field, which is also presents in both discourses, is the one of natural disasters. Since also left-wings-parities use this metaphor, it does not seem to be exclusively perceived as negative by the text producer. However, comparing people to flows or waves produce a negative effect, given that in this way they appear as unformed units neither controllable nor stoppable. This metaphor is also related to images of changes and movements. As waves, flows, and water constantly move, so do migrants. A natural event is, however, also defined by unpredictability, lack of control, and destructive potential. These attributes, not characteristic of migration, are shifted and inevitably linked ontologically to the migrant concept.

Metaphorical Field	German	Italian
Natural Disas- ter	Strom (Flüchtlingsstrom, Migrationsstrom): Kommt es zu einem grossen Strom von Asylsuchenden ist eine proportionale Aufteilung [] zu prüfen. CVP 18.03.2015, Massive Migrationsströme befinden sich [] auf dem Weg durch Europa. SVP 14.09.2015, Der Bruchteil eines Bruchteils dieses Flüchtlingsstroms wird bis Ende Jahr in der Schweiz um Asyl bitten. SP 09.09.2015	<i>Flusso</i> : Con flusso migranti forte rischio infiltrazioni terroristiche. PDL 16.02.2015, Tutti i nostri partner europei devono farsi carico del continuo flusso mi- gratorio. UDC 17.06.2015, Farci invadere da un flusso inarrestabile di profughi siriani o iracheni. Lega 19.10.2015
		Ondata: Alfano deve fermare subito l'ondata di sbarchi. Lega 09.04.2014, l'ondata di sbarchi che si è registrata negli ultimi tre giorni. PDL 14.04.2015, Tutte le volte che si presenta un'ondata migratoria come sta avvenendo [] dall'Egitto e dalla Siria. UDC 21.08.2013
Act of War		<i>invasione, invadere</i> : La nostra battaglia oltre che contro l'invasione è anche contro la costruzione di moschee. Lega 23.07.2015, L'invasione non ha fine, un'altra trentina di clandestini sono stati portati in alcuni comuni del comasco. Lega 20.02.2015, [una vera e propria invasione], [invasione di clan- destini]
Bible		<i>Esodo (biblico)</i> : C'è da aspettarsi che sarà un autentico esodo di massa. lega 19.05.14, An- cora oggi continua l'esodo biblico delle popolazioni del mediterraneo verso il nord. NCD 07.09.2015, Quella palesata in forme drammatiche dall'esodo biblico di profughi e migranti. SEL 16.09.2015

#### Table 21: Metaphors

Metaphors in the semantic field of war and religion are exclusively present in the Italian discourse. The first posses an absolutely negative connotation and actively aims to produce fear, presenting the landing of migrants as a territorial invasion. The semantic field of war also includes images of invading armies, burning cities, and genocides, which obviously evoke negative associations. The biblical metaphor has a less clear-cut position. On one hand, it is a reasonable comparison, describing entire populations forced to abandon their country, on the other hand, it contributes once more to the dehumanisation process. This metaphorical field is related to the emotive referential metonymies previously recognised in the Italian discourse.

By using all these metaphors to describe migrants text consumers actuate a predication strategy. In fact, the characteristics of the different metaphorical fields often associated with immigrants become ontological features of the latter.
We only detected two of the topoi related to emotive coercion presented by Hart (Hart [2010]), the topos of danger and the topos of diseases. The former is shared by both countries, even if its more developed in the Italian one. The only term identified among the Swiss-German list, used to implement the danger topos is *Terrorismus*. This, co-occurring with words such as *Terrorgefahr*, *Risikopotenzial*, and *Asylweg* is often correlated to the possibility of terrorist infiltrations among the applicants for asylum.

- (8.11) Dennoch fordert die FDP, aufgenommene Flüchtlinge auf ihr terroristisches Risikopotential hin zu überprüfen. Erfahrungen aus anderen europäischen Ländern haben gezeigt, dass auch über ein vom UNHCR zugeteiltes Flüchtlingskontingent Terroristen einreisen können. FDP 18.08.2015
- (8.12) Dies umso mehr, da immer häufiger Fälle von Einschleusung von Terroristen in unser Land über den Asylweg bekannt werden. SVP 28.06.2015

The enthymeme hidden behind affirmations like the ones in (8.11) and (8.12) works exactly like the ones explained by Wengeler (see section 2.2.2). Since some refugees might be terrorists, the country is in danger, and to avoid this danger asylum route and procedures should be stopped. In this case the context related rule of derivation effectively fulfil its role by transforming real premises in an unjustified conclusion as if it was the only logical one.

*Terrorismo* and *terrorista* appear also in the Italian list, in addition to many more terms that all reinforce the same danger topos. We summarized them all with their co-occurrences<sup>4</sup> and MWEs in Table 22.

<sup>&</sup>lt;sup>4</sup>Terms co-occurring with the investigated term are always reported between brackets.

Term	MWEs	Context
Terrorismo (non, paese, favoreggiamento)	[il terrorismo islamico]	Il terrorismo islamico non poteva che approfittare delle bib- liche migrazioni che hanno investito l'Europa, per mimetiz- zarsi infiltrarsi e organizzarsi. Lega 08.07.2014, C'è il ris- chio di una vera e propria emergenza terrorismo. Lega 07.07.2014
Rischio (infiltrazione, clandestino, terroristico)	[rischio di infiltrazioni ter- roristiche]	Forte rischio di nuovi terroristi islamici nel nostro territo- rio. Lega 15.09.2014, Con flusso migranti forte rischio in- filtrazioni terroristiche. PDL 16.02.2015
Pericolo		Che esporrebbe il nostro Paese al pericolo di infiltrazioni terroristiche. Lega 11.02.15
Infiltrazione (terroristico, rischio)		Ci sono rischi di infiltrazione di terroristi tra gli immigrati. PDL 22.01.2015
Invadere, Invasione	[farci invadere da], [inva- sione di clandestini], [una vera e propria invasione]	L'Isis minaccia di farci invadere da altri 500mila migranti? Lega 17.02.2015, Subito blocco navale, fermare invasione. PDL 07.06.2015
Islam, Islamico (culto, es- tremista)		L'islam e i clandestini si confermano un'insidia e un peri- colo che non ci possiamo permettere. Lega 26.06.2015, L'estremismo islamico ci sta accerchiando per colpirci al cuore. Lega 19.03.2015
Musulmano (cristiano, ji- hadisti)		Inaccettabile violenza musulmana Lega 15.05.15
Jihadisti (islam, islamico)		La minaccia terroristica jihadista e' alle porte del nostro Paese, solo duecento miglia separano la Sicilia dal nord della Libia []. PDL 16.02.2015, 800 jihadisti su coste li- biche pronti a farsi saltare in aria in Italia e in Europa. Lega 20.11.2014
Sicurezza (Italia, mare, immigrato)	[sicurezza e immi- grazione], [sicurezza dei nostri cittadini]	Su immigrati e sicurezza questo governo ha fallito. PDL 21.05.2015, Non possiamo aspettare che le nostre coste siano oggetto di continui sbarchi, che la Marina sia usata come taxi per clandestini, che le città siano invase di immigrati senza controllo e sicurezza per i cittadini. PDL 28.04.2015

Table 22: Italian Topos of Danger

As shown in the Table 22 many different terms are used to point out the same issue: migrants and refugees, more specifically, Muslim or Islamic ones, may possibly represent a terrorist threat for the country. ISIS is, in addition, possibly planning an invasion using those figures to implement it. To avoid this risk, all the landings must be stopped. This topos is far more developed than the Swiss-German one. Firstly, it involves many terms such as *rischio*, and *pericolo* that directly connect with the emotion of fear and consequently of anger for the unstable situation. Moreover, by mentioning the religious affiliation using *musulmano* and *islamico*, and using it insistently in relation to terrorism and the semantic field of danger, at a cognitive level, the automatic association between the Islamic religion and danger is established. It is the Islam itself that becomes a threat. This example shows how emotive and cognitive coercion are often dependent of each other.

The topos of disease is completely lacking from the Swiss-German discourse, but is fruitful in the Italian discourse.

Term	Context
Epidemia, Ebola, Virus (ebola, virus, Africa, misura)	Dopo Bologna, Varese e Gallarate, anche il Veneto registra i primi casi sospetti di ebola mentre le cronache quotidiane riportano di migliaia di clan- destini che vengono sbarcati nei porti del paese e smistati nei nostri territori senza nessun criterio e senza la certezza di avere adottato tutte le neces- sarie misure di sicurezza. Lega 05.09.2014, L'epidemia di Ebola rischia di espandersi anche in Sicilia e nel resto del Paese. PDL 09.11.2015, Il rischio epidemia è reale. Lega 07.06.2015
Scabbia, TBC (pomata, amuchina, malat- tia)	Ormai stiamo importando malattie scomparse da tempo in Italia. Lo sbarco di decine di persone con la scabbia a Salerno è un'ulteriore notizia inquietante. PDL 01.07.2015, Iniziano anche a trapelare le prime verità sui casi di scabbia in Lombardia tra i clandestini arrivati a Milano e a Monza. Lega 23.04.15, A Roma è scattato l'allarme scabbia e tbc perche' alcuni immigrati malati sono scappati dal centro di raccolta dandosi alla fuga. PDL 06.05.2015

Table 23: Italian Topos of Diseases

The examples reported in Table 23 show the actuation of the diseases topos. The fear of Ebola, Scabies, or TBC epidemic is connected with the landings of the migrants. The implied necessity to stop these arrivals appears from the contexts. The proclaimed health emergency generates fear in the text consumer, which perceives immigration as dangerous for its own safety. For example, despite the assurances of the Minister of Health about scabies that is not a threat any more<sup>5</sup>, this disease was still widely used in propagandistic ways to generate fear and negative feelings. Furthermore, underlining the resurgence of diseases extinct long time ago, implies a regress in the progress of society. Diseases extinct in Italy long time ago re-emerge to causes attributable to arrivals and landings. This generates anger in the text consumer.

Analysing different keywords we found out that the Italian discourse is characterised by an additional topos based on an emotion that we did not yet introduce: compassion. This feeling is usually associated to left-wing-party attitudes, which sustain immigration and reception policies. Instead, compassion is evoked and used by right-wing-parties (mainly Lega and PDL) to profess the closure of the borders and stop the receptionm in the topos of deadly immigration system. This topos states that the actual immigration systems is the cause of the death of thousand migrants that, as victims of people-smugglers, drown in the Mediterranean.

<sup>&</sup>lt;sup>5</sup>http://www.salute.gov.it/portale/salute/p1\_5.jsp?id=210&area=Malattie\_ infettive.

Term	MWE	Context
Criminale (ac- coglienza, profugo)	[operazione criminale Mare]	Un'operazione criminale il cui risultato è stato soltanto quello di molti- plicare il numero di immigrati clandestini provocando migliaia di morti in mare. Lega 03.10.2014, Ci chiediamo se queste persone sareb- bero mai partite se non ci fosse stata l'operazione criminale Mare Nostrum. Lega 15.09.2014
Mercante (morte, arric- chire)	[mercanti di morte]	Triton, nella sua versione 'potenziata', diventerà un grande Mare Nostrum Europeo, l'ennesimo favore agli scafisti e ai mercanti di morte. Lega 13.05.2015, Siamo di fatto complici dei mercanti di morte. PDL 21.04.2015
Carretta, Bar- cone (trappola)	[carrette del mare]	Altri 30 morti su un barcone. Lega 30.06.2014, Ben venga [] l' affondamento dei barconi. PDL 27.06.2015
Scafista (oper- azione, contro)	[aiutare gli scafisti], [contro gli scafisti]	Il governo spende milioni di euro per aiutare scafisti. Lega 22.07.2014, È necessario spostare immediatamente il baricen- tro dell'accoglienza dei profughi nel Nord Africa [] fermando gli scafisti, anche attraverso l' utilizzo delle forze armate. PDL 02.09.2015
Viaggio (ris- chioso)	[viaggi della morte ]	Incentiva i viaggi della morte con operazioni come Mare Nostrum e Triton. Lega 13.04.2015
Mare (nostrum, paese, Italia)	[morti in mare]	Mare Nostrum incoraggia scafisti. Lega 02.11.2015, Mare nostrum aveva due obiettivi: meno sbarchi e meno morti. I morti sono stati più di tremila, gli sbarchi quasi 150 Mila. Lega 03.10.2014
Pelle	[sulla pelle degli immigrati]	Basta business sulla pelle degli stranieri con i soldi dei cittadini. Lega 22.09.2015, Uno squallido e sporco business sulla pelle degli immi- grati. Lega 21.07.2015
Sangue (morto, coscienza, mani)	[sporche di sangue]	Prosegue scia di sangue del governo Renzi-Alfano. Lega 19.04.2015, Altri 180 IMMIGRATI MORTI grazie a Mare Nostrum, operazione di sangue. Lega 22.07.2014
Mediterraneo (cimitero, morto)	[morti nel Mediterraneo]	Le politiche di Renzi e Alfano stanno trasformando il Mediterraneo in un mare di sangue. Lega 16.04.2015, I barconi della morte non cesseranno mai di solcare le acque del Mediterraneo. PDL 23.04.2015 PDL
Tragedia (clan- destino, gov- erno, Renzi)		Centinaia di immigrati pronti a sbarcare e ancora una tragedia nel Canale di Sicilia. PDL 04.03.2015, Renzi, Alfano, Boldrini e Mogherini sono corresponsabili della tragedia in mare più grave di sempre. Lega 14.04.2015
Fermare (ac- coglienza, mare)	[fermare gli sbarchi]	Fermare le partenze, aiutarli a casa loro, subito! Le camicie di Renzi e Alfano sono sporche di sangue. Lega 30.06.2014, Per fermare gli sbarchi di clandestini e bloccare i mercanti di morte va ripresa la politica del Mediterraneo di Silvio Berlusconi. PDL 15.15.2014.

Table 24: Italian Topos of Death

In the sentences reported in Table 24 we found many terms used to describe the specific situation of migrants arriving by sea. The terms underline the critical conditions currently characterising the Mediterranean crossing. They refer to large boats loaded with migrants, journeys of death, and criminal network exploiting them. Migrants are described as victims and feelings of piety are evoked in the text consumer. Furthermore, the government is directly accused to be an accomplice to these never-ending sea tragedies. Metaphorical expressions, revealed by MWEs, such as *mani sporche di sangue* or *business sulla pelle degli immigrati* are very strong and suggestive at an emotive level. The first metaphoric expression implies that the government exponents have committed the killing personally. The second one, in the same manner, states that the system is exploiting migrants.

The Italian operation *Mare Nostrum*, for example, which began on October 2010 by the government Letta (16.11.13 - 10.12.2013), is one of the main targeted actions by government opponents. The operation, established to tackle the increased immigration to Europe, permitted to at least 150'000 migrants to safely arrive to  $Europe^{67}$ . However, it was considered as part of the corrupted and criminal immigration business, encouraging migrants departures and contributing to the deadly system.

The whole immigration management system is understood as corrupted, implemented to enrich the criminal organisation behind it, and causing migrants to die. Instead of proposing the introduction of humanitarian aid corridors to save migrants, right-wing-parties strongly suggest to stop the landings and close the borders. This should be perpetuated in every way possible, either with a military intervention in Libya or by sinking the boats directly on the North African coasts. Taking into account this conclusion, we can easily understand that the compassion for the fate of migrants is only an argumentative expedient. This emotion is evoked to justify Italian-centric solutions which do not take the migrants problems into consideration at all.

The term *Corridoi* figures in the Italian list, co-occurring with terms such as *umanitario*, *attivazione*, *sicuro*, *catastrofe* and recurring in the MWE [di corridoi umanitari sicuri]. Examples (8.13) and (8.14) report the keyword in its contexts. A humanitarian corridor consists in a temporary demilitarised zone that allows the exit of refugees from a crisis region, or the safe transit of humanitarian aid. In

<sup>&</sup>lt;sup>6</sup>Mare Nostrum Operation. Ministry of Defence of Italy :http://www.marina.difesa.it/EN/ operations/Pagine/MareNostrum.aspxRetrieved 16 April 2015.

<sup>&</sup>lt;sup>7</sup>In October 2014 the operation was suspended and substituted by the European operation *Triton. Triton* focuses, unlike *Mare Nostrum*, on border protection rather than search and rescue. Furthermore, it operates closer to Italy covering exclusively up to 30 miles from the coast.

our case it would consist in the institution of safe passages, by boat or plane, for migrants departing from critical areas: this would allow them to arrive in Italy and submit a request for asylum without having to cross the Mediterranean transported by smugglers<sup>8</sup>.

- (8.13) Serve creare corridoi umanitari a difesa dei civili. SEL 09.10.2014
- (8.14) La creazione di corridoi umanitari sicuri, sia via terra che via mare, per permettere a chi scappa da guerra, fame e povertà un approdo sicuro. Solo così L'Europa sarà all'altezza dei suoi valori e si potranno salvare migliaia di vite umane. SEL 06.11.2015

The proposal regarding the introduction of humanitarian corridors is exclusively effectuated by SEL.

As shown by the analysis of both metaphorical fields, used to actuate the predication strategy, and topoi of danger, diseases, and death, the domain of emotive coercion is fertile for the Italian discourse: many argumentative strategies are implemented to achieve this kind of coercion in the text consumer. The Swiss-German discourse, on the contrary, does not shows many references to this kind of strategy, leaving room, as we are going to show, to many implementations of the cognitive coercion.

#### 8.3.3 Cognitive Coercion

The last argumentative strategy left to analyse is cognitive coercion. Therefore, we will take into consideration all the topoi of this field.

From different lemmas of the Swiss-German keywords-list we recognised the topos of economical utility. This affirms that since an action produces or does not produce an economical benefit for the ingroup, it should be actuated or not actuated. This topos in turn includes two more topoi, one for the positive and one for the negative case:

- Topos of financial benefits: the outgroup member brings economical benefits to the ingroup;
- Topos of disadvantage: the outgroup member does not produce any benefit

<sup>&</sup>lt;sup>8</sup>From January 2016 Italy opened the first humanitarian corridor for refugees in Europe. The project, designed and implemented by the *Community of Sant'Egidio* and the *Federation of Protestant Churches in Italy (FCEI)*, provided the emission of thousand visas to join Italy safely and submit the asylum requests. http://www.santegidio.org/pageID/1165/langID/ it/itemID/756/SCHEDA-Cosa-sono-i-corridoi-umanitari.html.

#### for the ingroup;

Both topoi are present in the Swiss-German discourse but lacking in the Italian one. Whereas the financial benefit topos refers to the European intern immigration, in the topos of disadvantage both sub-discourses (European migrants crisis, European intern immigration) previously identified interweave.

Term	MWEs	Context
Königsweg (bilateral, Freihandelsabkommen, Erfolgsweg)		Die FDP ist die einzige Partei, die nie an den bilateralen Verträgen gezweifelt hat. Sie sind der Königsweg für die Schweiz. FDP 27.06.2015, Den bilateralen Königsweg kon- sequent weiterführen. FDP 17.03.2014
Fachkräfteman- gel, Fachkraft (Fachkräftezuwan- derung, Schweizweit)	[Fachkräfte aus dem Aus- land]	Viele Branchen kämpfen mit einem Fachkräftemangel, der sich nach der Annahme der Initiative gegen Masseneinwan- derung wohl noch verstärken wird. FDP 19.10.2014, Die CVP fordert weiterhin mehr Engagement im Kampf gegen den Fachkräftemangel in der Schweiz. CVP 26.03.2014

Table 25: Swiss-German	Topos of Benefits
------------------------	-------------------

Term	MWEs	Context
Renitent (Asilbewerber, Asylsuchende)	[kriminelle und renitente Asylbewerber]	Ein entschiedenes Vorgehen fordert die SVP auch gegen kriminelle und renitente Asylbewerber. SVP 31.08.2011, Ebenfalls einstimmig beschlossen hat die FDP-Liberale Fraktion die Dringlichkeit für die [] besondere Unter- bringung von renitenten Asylsuchenden. FDP 12.06.2012
Hochqualifisiert, Schönfarberei (Aus- sage, Bildungsabschluss, Maturitätsquote, Mär)		Die bereits zum Mythos stilisierte Aussage, dass vornehm- lich hochqualifizierte Einwanderer in die Schweiz kommen wird erneut aufgetischt. SVP 11.06.2013, Die Behauptung, es wanderten nur Hochqualifizierte ein ist ebenso falsch wie die Mär von neuen Arbeitsplätzen als Folge der Zuwan- derung. SVP 03.05.2013, Einmal mehr wurde auch die Mär von der hochqualifizierten Zuwanderung mit Blick auf die Bildungsabschlüsse der Zuwanderer aufgetischt. SVP 06.12.2013

Table 26: Swiss-German Topos of Disadvantage

As we can see from Table 25 and 26 the two topoi are interconnected regarding the topic of European immigration. While the first underlines the importance of foreign skilled employees, stating difficulties to recruit sufficient Swiss workers in determinate sectors, the second one describes this claim as unreal and as a fairy-tale. The fact that mainly highly qualified people arrive in Switzerland is doubted and rejected, taking into consideration statements regarding the qualifications of migrants. The political parties distribution in this case is clear: liberal and Europe-oriented parties like FDP and CVP implement the benefit topos, while the SVP, nationalist party, actuate the disadvantage one. Asylum seekers are instead exclusively took into account to fulfil the disadvantage topos. Some of them are described as *renitent*, or wittingly not willing to integrate and to actively contribute to the ingroup community. This frequent juxtaposition, which contributes to the stereotype of the idler migrant, does not take into consideration the legal impossibility of working during the asylum decision waiting period. Regarding the asylum seekers issue also the opposition in the party distribution previously mentioned ends. Not just SVP, but also FDP, and CVP built and implement this topos. This contributes to the opposition between useful and unnecessary migrants described by Gould (Gould [2006]:171)<sup>9</sup>.

On the contrary, the Italian discourse distinguished itself regarding the financial field for the topoi of burden finance and displacement. The former consists in reporting migrants in monetary terms, as costing a specific price to the ingroup, the latter describes members of the outgroup having access to special socio-economic resources negated to the ingroup. The difference with the topos of disadvantage consists in the fact that not only immigration does not contribute to the common good, but is also highly expensive for the ingroup.

The topos of burden on public finance is attested in the Italian discourse by the verb *spendere*, which co-occurs with terms such as *miliardo*, *Euro*, and *soldo*.

- (8.15) Ormai la misura è colma e prima di spendere altri soldi e trovare nuove strutture d'accoglienza, la provincia di Bolzano dovrebbe preoccuparsi di investire le risorse a disposizione per creare nuovi posti di lavoro e fornire abitazioni ai propri cittadini rimasti senza casa. Lega 04.11.2014
- (8.16) I milioni di euro che si vogliono spendere per ospitare chi è palesemente clandestino e oggi si finge profugo. Lega 14.04.2015
- (8.17) Non è questo il modo di aiutarli, spendere 35 euro al giorno per queste persone è follia. Lega, 15.05.2014

Examples (8.15) - (8.17) show how the migrant reception process is reported exclusively in monetary terms, considering it as a waste of public money that could be employed to support the ingroup members. At a cognitive level migrant are not perceived as persons in need but rather as monetary loss.

The topos of displacement, is structured around more terms, often recurring in crystallised MWEs.

<sup>&</sup>lt;sup>9</sup>We are not going to deepen this opposition in this contribution.

Terms	MWEs	Context
Vitto, Alloggio (sigaretta, cellulare, abbigliamento, paradiso)	[vitto e alloggio]	È inaccettabile che lo Stato paghi vitto e al- loggio ai profughi mentre lascia per strada e senza lavoro i nostri cittadini. Lega 14.04.2015, È questa la stella polare che guida la sinistra anche nel campo dell' immigrazione: vi porti- amo in Italia e vi diamo vitto e alloggio. PDL 13.04.2014
Hotel, Albergo (allog- giare, stella, lusso, sistemazione)	[hotel di lusso], [alberghi di lusso]	Clandestini alloggiati in centralissimi hotel pluristellati. Lega 23.09.2014, Bus dedicati, sigarette, alloggi di lusso, schede telefoniche, vitto e alloggio pagato: l'Italia [] è diventata il paradiso dei clandestini e l'inferno per gli ital- iani Lega 18.06.2015

#### Table 27: Italian Topos of Displacement

Board and lodging, as well as luxury hotel, and telephone cards, have become fixed expressions in the Italian discourse about immigration. The claims that migrants have free access to these goods, often unproven and overstated, constitute the topos of displacement (Table 27). The text consumer is confronted with the representation of an outgroup enjoying advantages that are denied to him. A particular cognitive process is here invoked: the overturning of the ingroup - outgroup dichotomy. The traditional representation, that pictures the ingroup members as belonging to a stable and democratic country and the outgroup members on need of help, is completely reversed: while the ingroup faces a financial crisis with sacrifice and suffering, the outgroup lives in luxury. This process leads to the conviction that migrants enjoy special privileges at expenses of the ingroup. At a cognitive level migrants become the scapegoats to blame for all the socio-economical problems of the ingroup.

To straighten this reversed representation we identified the topos of inverse racism (Table 28), again only recurring in the Italian discourse. This consists in the identification - carried out by right-wing populist parties - of a reversed racism perpetuated against Italian citizens by government exponents and left do-gooders. Right-wing parties, traditionally associated with racism and often accused of it in the public discourse, appropriate this concept and use it for their cause.

Terms	Context
Razzismo, Razzista (xenofobia, non, rovescio)	Italiani stufi di razzismo alla rovescia 02.09.2015. PDL 24.04.2015, I clandestini chiedono e ottengono, mentre gli italiani fanno la fame: basta razzismo, Per dire basta a questo razzismo al contrario. Lega 08.04.2015. Condanniamo senza se e senza ma la scelta razzista di agevolare unicamente gli immigrati, voltando le spalle ai nostri ragazzi. Lega 14.07.2015
Buonismo, Buonista (ipocrita, emotivo, fin- gere)	Questo falso buonismo legato alle politiche sconsiderate di alcuni ministri porterà l'invasione nel nostro Paese. Lega 21.01.2014, Buonismo 'sinistro' incentiva invasione. Lega 10.04.2014

Table 28: Italian Topos of Inverse Racism

Government exponents, left politicians and voters, are accused by Lega and PDL to be racist against Italians.

(8.18) Racism: any action, practice, or belief that reflects the racial world-view that humans are divided into separate and exclusive biological entities called "races," that there is a causal link between inherited physical traits and traits of personality, intellect, morality, and other cultural behavioral features, and that some races are innately superior to others<sup>10</sup>.

Following the definition at example 8.18 the inappropriate use of the racism concept in this context is clear. All the requirements to talk about racism are lacking as the race notion has nothing to do with this specific situation. However, this do not prevent the topos to be very effective. It, in fact, builds upon the topos of displacement and gives rise to a strengthening of the inverse dichotomy where the Italian citizen is a discriminated victim, and the migrant enjoys all privileges. The construction of this topos combines with the Italian trend of mistreating the negative face already discussed. The result is an extremely negative representation of the ingroup situation, limited in its actions and freedom, that greatly reinforces the contrast with the condition of the migrants, described as privileged.

As already mentioned the topos of exploitation is the most developed in the Swiss-German discourse. This topos activates in the text-consumer the cheaterdetection modus discussed by Hart (Hart [2010], see section 2.2.2.3). In fact, migrants are represented as social cheaters that unfairly benefit from the ingroup humanity and generosity.

<sup>&</sup>lt;sup>10</sup> "racism", Britannica, http://www.britannica.com/topic/racism. Retrieved 03.06.2016.

Term	MWEs	Context
Humanitäre (Tradition)	[die humanitäre Tradition der Schweiz ]	Die Schweiz hat eine lange humanitäre Tradition auf die wir stolz sind. CVP 03.04.2011, Unsere humanitäre Tradi- tion und unsere Gastfreundschaft wer- den immer schamloser ausgenutzt. SVP 09.12.2011
Scheinasylant (echte, Flüchtlinge)		Scheinasylanten und Kriminelle strömen massenhaft in die viel zu attraktive Schweiz. SVP 15.05.2013, Anstatt die Missbrauche zu bekampfen, lasst sie es zu, [] dass die Attraktivitat der Schweiz fur illegale Scheinasylanten gesteigert wird. SVP 14.01.2016
Leib (Leben, Bedroht, Sozialmigranten)	[nicht an Leib und Leben ], [an Leib und Leben bedroht]	Wer tatsächlich an Leib und Leben bedroht ist, hat in der Schweiz nach wie vor An- recht auf Asyl. CVP 18.06.2012, Die Schweiz hat eine lange Tradition der Auf- nahme von Flüchtlingen, die an Leib und Leben bedroht sind. SVP 13.10.2015, Den Flüchtlingsstatus sollen jene erhalten, die auch tatsächlich an Leib und Leben bedroht sind. FDP 30.06.2015
Schutzbedürftig		Falls jemand nicht schutzbedürftig ist, muss dies zum Verlust der Aufenthalts- bewilligung in der Schweiz führen. FDP 30.06.2015
Asylmissbrauch, Aslymissbrauchler		Die Schweizer Asylpolitik ist heute bes- timmt von Asylrechtsmissbrauch, von Ab- surditäten und der Untätigkeit sowie der Führungslosigkeit der Verantwortlichen im Asylbereich. SVP 05.06.2012
Asylchaos (Asylzentrum Sommaruga)		Die SVP verlangt vom Bundesrat, dass er endlich wirkungsvolle Schritte gegen das vorherrschende Asylchaos einleitet. SVP 03.09.2011, Asylchaos stoppen, Verfahren beschleunigen. FDP, 05.06.2011
Asylindustrie (profitieren, kostenlos Asylverfahren)		Eine ganze Asylindustrie profitiert vom Asylmissbrauchswesen in unserem Land. SVP 09.06.2013, Profitiert von dieser unhaltbaren Situation haben die Wirtschaftsmigranten und die Asylindus- trie. SVP 09.06.2015
Asylproblem, Asylprob- lematik (reorganisation, verwalten)	[Problemen in Asylbereich]	Der Bundesrat muss nun endlich die Lösung der Asylprobleme im Inland ange- hen und die Attraktivität der Schweiz als Asyldestination senken. SVP 14.11.2014

Table 29: Swiss-German Topos of Exploitation

This topos is based on the Swiss-German positive face addressing strategy since it is developed starting from the representation of a helpful and humanitarian Switzerland. The sequence of relevant keywords reported in Table 29 shows clearly the constitution of this topos. Switzerland posses a long humanitarian tradition and distinguishes itself through the culture of hospitality. Some migrants, the so-called bogus asylum seekers, abuse this culture and tradition. German language posses, as we saw talking about metonymical referential strategies, many compounds to address these individuals. These terms distinguish bogus asylum seekers, that are not risking their lives and take advantage from the system (nicht an Leib und Leben bedroht), from the real people in need. We are once again faced with a dichotomous representation of individuals recognised as in need of protection and individuals labelled as bogus. Furthermore, the Swiss current asylum system is described as an industry, where only economic migrants (bogus asylum seekers) take advantage of the common good, preventing the help to real asylum seekers. Switzerland is presented as an exploited country despite its humanitarian tradition, already broadly implemented. The only solution to this social and economical injustice consists in the reorganisation of the whole asylum system and the tightening of the to generous asylum law. This topos causes the cognitive automatic association between the migrant and the cheater/exploiter.

Exploitation appears in the Italian discourse too, but is far less developed. This is manifested mainly in the contrastive representation of alleged and real refugees. Terms such as *sedicente* and *presunto* call into question the recognition of refugee status for certain individuals. Both share *profugo* as the highest probable term with which they co-occur, and *presunto* recurs in the following MWE [i presunti profughi].

- (8.19) Ogni anno 15mila presunti profughi vengono rispediti qui dagli altri Paesi dell'Unione Europea. Lega 04.03.2015
- (8.20) A vicenza non c'e' posto per altri 40 presunti profughi. Lega 22.03.2014
- (8.21) l'immigrazione clandestina in continua crescita anche per le continue politiche di accoglienza di sedicenti profughi. Lega 12.11.2014
- (8.22) L'ultimo naufragio ha provocato almeno 40 morti, mentre a Taranto sono sbarcati centinaia di altri sedicenti siriani tra le proteste dei residenti per le strutture al collasso. PDL 11.05.2015

Sedicente co-occurs also with *siriani*, implying that also the belonging to the Syrian nation is doubted (see example 8.22). While Syrians are more easily recognised as rightful asylum seekers, mainly migrants arriving from Africa are considered to be cheater. The exploitation issue is implied in the Italian discourse, but not in

an explicit way.

Additional topoi in the Swiss-German discourse, aiming to negatively influence the cognitive representation of the migrant, actuating the predication strategy, are the ones of culture and crime. These follow the trend described until now, supplying a stable base for the establishment of the previously discussed topoi of disadvantage and exploitation (see section 2.2.2.3).

Term	Context
Ausländerkriminalität (Schengen, Grenzübergreifend, Asylmissbrauch)	Zur Bekämpfung der zunehmenden Ausländerkriminalität []. SVP 30.05.2012, Die hohe Ausländerkriminalität ist einer der Hauptgründe für die zunehmenden Spannungen zwischen Schweiz- ern und der ausländischen Wohnbevölkerung. CVP 11.11.2015
Kriminell (Ausländer, Ausschaffung, Schweiz)	Wir wollen die kriminellen Ausländer ausschaffen. SVP 07.10.2011, Es braucht nun endlich griffige Massnahmen gegen kriminelle Ausländer. SVP 28.12.12, Mit anderen Worten: Es werden im- mer mehr Ausländer mit kriminellem Potenzial eingebürgert. SVP 26.09.07
Ausweisen (straffällig, Drogendelikt, Ille- gale)	Eine Steuerung der Zuwanderung und eine konsequente Ausweisung krimineller Ausländer sind nun umgehend umzusetzen. SVP 13.09.2015, Kriminelle vor Ausweisung geschützt. SVP 19.03.2015

Table 30: Swiss-German Topos of Crime

As shown from the few examples reported in Table 30 crime is often associated with the migrant figure in the Swiss-German discourse. The crime rate among foreigners is described as a growing phenomenon against which strict measures are needed. This fits perfectly in the classic stereotypical framework that associate only negative qualities to migrants. The two popular initiatives  $F\ddot{u}r$  die Ausschaffung krimineller Ausländer (Ausschaffungsinitiative) and Zur Durchsetzung der Ausschaffung krimineller Ausländer (Durchsetzungsinitiative), proposed by the SVP respectively in November 2010, and February 2016, show perfectly the frequent bonding between the concepts of foreigner and criminality<sup>11</sup>. Criminality is so often associated to foreigner that it became almost a natural attribute of the whole category.

<sup>&</sup>lt;sup>11</sup>Both *Durchsetzung* and *Ausschaffung* resulted in the Swiss-German keywords-list.

Term	Context
Sprache (kultur, lernen, integration)	(Zuwanderer, die sich weigern, die Sprache zu erlernen, sich nicht an die Schweizer Rechtsordnung [] sind nicht integrationswillig und haben in der Schweiz nichts verloren. SVP 08.08.2013, Viele Einge- wanderte lernen unsere Sprache nicht und halten sich nicht an un- sere Regeln. SVP 27.07.2010
Rechtsordnung (stützen, Identitätskon- trolle)	Deshalb müssen wir fordern, dass Einwanderer unsere Rechtsor- dnung integral zu akzeptieren haben. SVP 12.2012, Wer in die Schweiz kommt, hat sich an die Rechtsordnung zu halten, sich anzu- passen []. SVP 19.03.2015
Identität (Veränderung, Wandel)	Damit die Schweiz ihre positive Identität wahren kann, brauchen wir [] eine vernünftige Zuwanderungspolitik gegen aussen. SVP 20.03.2014, Wir dürfen unsere Identität nicht aufgeben. SVP 13.10.2015

Table 31: Swiss-German Topos of Culture

Table 31 contains the keywords that compose the topos of culture. Concepts such as language and legal order are strongly related to national identity. Migrants are often described as not willing to integrate in the country, refusing to learn the local language and breaking the country's laws. This is perceived as harmful and dangerous for the national identity. Furthermore, the failed integration is described as intentional: despite the attempts of the ingroup, outgroup members refuse to cooperate. This lays the foundation for a description that paints the migrant as most likely to commit criminal actions since they are not part of the community. Consequently, they are also more likely to exploit the social system.

Finally, the last topos recognised as specific of the Swiss-German discourse is the topos of overload. This posses the particularity of combining the two big sub-discourses about immigration, European migrant crisis and European intern immigration, which usually are treated separately. In fact, as we saw in the previous examples, most of the topoi regarded exclusively the topos of European migrant crisis. The concept of overload is particularly strong: stopping the access of migrants to the country is not represented as a choice but as a necessity. The country is compared to a bucket full of water with no capability of containing one further drop. In the Swiss-German discourse it is stated that since the country system is overloaded of foreigners, employees or refugees, the independence from the EU needs to be established again, in order to act against this overload.

Term	Context
Belasten	Zuwanderer aus der EU verdrängen Arbeitnehmer aus Drittstaaten, die wiederum nicht in ihre Heimatländer zurückkehren und die Sozialwerke belasten. SVP 23.05.2011, überfüllte Züge, ver- stopfte Strassen, steigende Mieten und Bodenpreise, Verlust von wertvollem Kulturland, Lohndruck, Ausländerkriminalität, erhöhte Ar- beitslosigkeit in den Branchen mit hoher Zuwanderung, Asylmiss- brauch, belastend hohe Ausländeranteile in der Fürsorge und an- deren Sozialwerken. SVP 27.09.2013
Unsteuerbar (Einwohnern, dramatisch, ex- plodieren)	Die Zuwanderung aus der EU unsteuerbar geworden ist und das Abkommen über die Personenfreizügigkeit dringend angepasst wer- den muss. SVP 10.10.2011, Mit der Einführung der Personen- freizügigkeit ist die Situation auch in diesem Bereich unsteuerbar geworden. 05.03.2012, SVP
Arbeitslos (Gesamtarbeitslösigkeit, attrak- tiv)	Fast 77% der neu arbeitslos gewordenen Personen sind Ausländer. Und davon stammen wiederum 76% aus der EU. SVP 09.12.2013, Heute sind zirka 80 Prozent der anerkannten Flüchtlinge im er- werbsfähigen Alter arbeitslos und beziehen Sozialhilfe. CVP, 07.11.2014
Arbeitsmarkt (Schweiz, EU, Land)	Insbesondere in wirtschaftlich schlechten Zeiten muss die Schweiz den Zugang von Ausländern zu unserem Arbeitsmarkt und zu den Sozialleistungen regulieren können. SVP 11.05.2009, Spielraum bei der Personenfreizügigkeit nutzen und negative Auswirkungen auf Wohn- und Arbeitsmarkt begrenzen EVP 27.09.2011
Eigenständig (Steuerung, Zuwanderung, EU)	Die Zuwanderung soll wieder eigenständig steuerbar und be- grenzbar werden. SVP 15.06.2011, Die Schweiz steuert die Zuwan- derung eigenständig. SVP 23.05.2014

Table 32: Swiss-German Topos of Overload

Examples of Table 32 show clearly the development of the topos: the social system is overloaded because of refugees and European unemployed. The employment market, but also the country borders, need to be controlled exclusively by Switzerland, without any influences or pressures from the European Union. The solution is independence from the European Union. Consequently, it is necessary to perform a revision of the bilateral agreements. Also in this case the arguments, which constitute the topos, are developed by the SVP.

### 8.4 Observations

The analysis of the extracted keywords, co-occurrences, and MWEs in their original context allowed us to investigate differences and similarities between the Swiss-German and the Italian political discourse about immigration. In both cases the results showed that the implementation of argumentative strategies is performed, above all, by right-wing-parties. In recent years, radical right parties have reemerged as electoral forces in Europe (Appel [2012]). In a time of general crisis, these nationalist and populist movements base their election campaigns on problems related to the management and reception of migrants. Therefore, we recognised a general trend that pushes for an immigration reduction and control, common to both countries. Furthermore, the ingroup - outgroup differentiation is strongly present in both discourses and consists in the base according to which the discourses are structured. However, a differing setting and development of the argumentation has been clearly recognised.

The Swiss-German discourse is structured around the key topics of finance and economy. As Gould claims, "business and the resulting prosperity are the principal and common characteristics of the confederation and the nation" (Gould [2006]:166). Consequently, immigration has always been compared to an economic issue in Switzerland (Kueng [2005]:1). Despite the internal controversial regarding EU-immigration emerging from the topos of economical utility, that opposes financial benefits and disadvantages, the immigration topic is constructed around the exploitation topos and the activation of the cheater-detection modus. In fact, it corresponds to the central argumentative strategy of the whole Swiss-German immigration discourse. Almost all the other identified analysis categories somehow contribute to it, either by creating the basis of the exploitation topos or by sustaining it. The positive face addressing strategy creates the conditions for the representation of an humanitarian and generous Switzerland being exploited. The constant reference to a successful model, characterised by prosperity and wealth, leads to diverse effects. While, on the one hand, it reinforces the image of a just and democratic Switzerland opposed to criminal exploiters, on the other hand, it describes the country as attractive and plausible for exploitation. Outgroup members are painted exclusively in relation to the ingroup economical exploitation and not using more general attributes as in the Italian case. Referential metonymies, for example, are mostly compounds such as Scheinasylant, Scheinflüchtling, Wirtschaft-, or Sozialmigrant, which all imply doubtful motivations for immigration. As we noticed emotive coercion does not prominently figure in the discourse despite the presence of the metaphorical field of natural disaster used for describing migrants, or the topos of danger. The general strategy consists much more in fixing at a cognitive level the opposed representation of an ingroup: generous, correct, and exploited, and an outgroup: criminal and exploiter. Also the analysed topoi of culture and crime aim at achieve precisely that objective. The action of representing the migrant as non-cooperative and not willing to integrate despite the possibilities, makes it the perfect candidate to commit crimes and exploit intentionally the Swiss system.

On the contrary, the Italian discourse is primarily characterised by the evoking of emotions. These, are aroused by focusing exclusively on the problems related to immigration using a language strictly linked to fear, crimes, pain, and degradation. The topoi of danger and diseases, for examples, are extremely developed and describe a current situation that appears on the verge of an imminent catastrophe. Migrants become bearers of risks and problems at sanitary and security level. Emotions traditionally related to the immigration topic, such as anger and fear are, furthermore, induced in the text consumer by the use of both referential metonymies and predication metaphors. Outgroup members referenced through negative attributes (such as illegality, desperation, otherness) lose their human condition. These negative attributes are, furthermore, generalised and perceived as ontologically belonging to the migrants category. The metaphorical fields of natural disasters, war, and biblical references contribute even more to the dehumanisation process: migrants are no longer single individuals but appear as an uncontrollable and harmful mass. These masses, composed of water (flows, waves) or persons (invaders, exodus), threaten the host country as well as its inhabitants. In addition to fear and anger we identified also the evoking of compassion, emotion not traditionally used in this context. Compassion is aroused in relation to the terrible conditions of migrants, implying that the current immigration system causes migrants death. Compassion is, however, not used to claim humanitarian corridors but to close the borders and stop the landings.

In the Italian discourse also strategies performing at cognitive level are influenced by emotions. The traditional dichotomous relationship between ingroup and outgroup is diametrically reversed causing anger in the text consumer. Italian citizens are, indeed, described as victims of displacement and inverse racism. While Italians pass through an economical crisis, suffering hunger and poverty, migrants live as privileged inhabitants surrounded by luxury at expenses of the citizens (topos of displacement). Italians are not racists, but victims of racism actuated by a government that discriminate them (topos of reversed racism). The actuation of all these strategies is supported by the mistreating of the text consumers negative face. The upturned conception of the ingroup - outgroup is much more efficient in a context where the ingroup members are continuously reminded of the multiple disadvantages, restrictions, and deprivations, to which they are subject.

## 9 Conclusion

As we pointed out in the introduction, the research questions of this project are carried out at two different levels. On one hand, they were related to the proposed method, a corpus-driven approach to analysing discourses, on the other hand, they were concerned to the analysed topic: immigration in the political discourses of German-speaking Switzerland and Italy.

In light of the proposed analysis it becomes clear that the utilisation of a corpus-driven CL approach to statistically calculate discourses is considerably productive for enriching traditional DA. This is mainly due to the advantage provided by such approach: the possibility of working inductively. By retrieving topic-significant elements, especially linguistic patterns, we gained a solid basis for formulating hypotheses from real data. Linguistic patterns offer, moreover, the possibility to include, beside the main topics and contents, also the manner in which the discourse is constructed at the linguistic level (Bubenhofer [2008]:431). The corpus is not used to verify a pre-existing hypothesis about the discourse, but rather to provide them. In this perspective it is possible to eliminate pre-formulated hypotheses and, moreover, to notice linguistic patterns that we would otherwise miss. Of course hypothesis suggested by the corpus have then to be analysed with the traditional methods of DA in order to be confirmed. The alternation between the corpus-driven and the corpus-based approach allows, therefore, the implementation of an analysis that is both quantitative and qualitative. This is proven to be efficient in order to recognise linguistic patterns and to re-conduce them to traditional DA categories. Metaphors, topoi, and metonymies are all elements that embody the argumentative strategies, which compose the essence of discourses.

This approach allows, moreover, to obtain concrete linguistic elements in two different languages and to compare them with the aid of SMT. The addition of SMT allows the implementation of a direct multilingual translation system, which leads to a concrete comparison of the two discourses. The use of SMT implemented in a systematic manner, by directly comparing the significant keywords previously extracted, provides a solid basis for the analysis. However, the presence of keywords in both lists does not automatically prove similarity: keywords need to be analysed in their context to understand if they are used in the same way in both discourses. The addition of co-occurrences and MWEs is extremely useful in this regard, integrating the analysis with further information. The introduction of multilingualism is definitely significant, particularly in the context of a topic that transcends national borders, such as immigration.

The combination of these strategies led us to recognise different settings behind the argumentative approaches regulating the political discourses of the German speaking part of Switzerland and Italy regarding immigration. We recognised similarities, but especially differences in the implementation of coercion, a strategy typical of politics. Coercion, a practice used to force someone to act in a manner by implementing some sort of persuasion, is usually emotive or cognitive in this context. The Italian discourse focuses mostly on emotive coercion, evoking fear, anger, and false compassion in relation to arrival by sea, reception, and management of the current immigration issue. Even the implementation of cognitive coercion is influenced by emotions in the Italian discourse. This is characterised by a reversed representation of the traditional dichotomy that opposes a we part of the ingroup to a *them* part of the outgroup. The outgroup member is no longer the victim in need of help, but becomes the privileged person who enjoys wealth, luxury, and social protection at expenses of the ingroup. The latter, on the contrary, is represented as the disadvantaged, victim of racism perpetuated by its own government. The Swiss-German discourse, on the other hand, shows a large implementation of cognitive coercion aiming to activate the cheater-detection modus. The discourse is composed by different argumentative structures, almost all of them aiming to influence the text consumer representation of the outgroup member in relation to economical exploitation. The host country, described as generous and helpful, with a long humanitarian tradition, is presented as exploited by migrants, which abuse the current system.

### 9.1 Limits and Future Prospects

As shown in this project, the combination of corpus-driven approach with SMT opens up prospects of future researches with a number of improvements possibilities. We decided to limit our approach to single-unit translation, working exclusively with translated words. This procedure allowed us to compare directly only the extracted keywords, from which we started to perform the analysis. This brings some limitations, particularly in the comparison of structurally different languages such as German and Italian. As we pointed out several times, whereas German is a fruitful language for the creation of compound words, Italian is not. This obviously leads to imbalances in the one-to-one matching and translation of terms.

- (9.1) asylsuchende : richiedente
- (9.2) freizügigkeitsabkommen : circolazione

Examples (9.1) and (9.2) clearly show the problem: Asylsuchende should be translated with rechiedente di asilo and Freizügigkeitsabkommen with accordi di libera circolazione. We partially solved this question by splitting German compounds, not covering, however, every problematic case. Furthermore, co-occurrences and MWEs can not be compared with this method, and were used only in a further step to enrich the subsequent analysis. Working with translated phrases or chunks in addition to single words would provide a solution to this problem. German compounds could thereby be directly translated and matched with the respective Italian expression. Moreover, crystallized patterns could also directly be matched and translated. Enriching the word-based approach with a phrase-based one would improve the procedure presented in the project as a whole, leading the comparison to another level.

We choose a general and vast topic about which we supplied an analysis capable of capturing the main differing elements in argumentation, but certainly nonexhaustive. The selection of more specific sub-topics in the filed of immigration could lead to a more precise and deep analysis. We also overlooked linguistic entities such as discourse style, expressed by the selection of a determinate variant under its paradigmatic alternatives (Bubenhofer [2009]:56), focusing on the most evident argumentation patterns. Furthermore, the consideration of stylistic factors would be another way to enrich the analysis. Finally, the addition of the investigation level of diachrony or of party political differences would be challenging: the presented method could surely prove himself fruitful for such analysis.

## References

- *Glossar 2.0 zu Asyl und Migration*. Europäische Kommission, 2012a. Europäischen Migrationswerknetz EMN.
- *Glossario 2.0 sull'asilo e la migrazione*. Commissione Europea, 2012b. Rete Europea sulle Migrazioni EMN.
- D. Antelmi. *Comunicazione e Analisi del Discorso*. De Agostini scuola, Novara, 2012.
- M. Appel. Anti-immigrant propaganda by radical right parties and the intellectual performance of adolescents. *Political Psychology*, 33(4):483 493, 2012.
- G. Brown and G. Yule. *Discourse Analysis*. Cambridge University Press, Cambridge, 1983.
- N. Bubenhofer. Diskurse berechnen? Wege zu einer korpuslinguistischen Diskursanalyse. In J. Spitzmüller and I. H. Warnke, editors, Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene, Linguistik – Impulse & Tendenzen, pages 407–434. de Gruyter, Berlin/New York, 2008.
- N. Bubenhofer. Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskursund Kulturanalyse. Sprache und Wissen 4. De Gruyter, 2009.
- N. Bubenhofer and J. Scharoth. Korpuslinguistische Diskursanalyse: Der Nutzen empirisch-quantitativer Verfahren. In U. H. Meinhof, M. Reisigl, and I. H. Warnke, editors, *Diskurslinguistik im Spannungsfeld von Deskription und Kritik*, Diskursmuster - Discourse Patterns 1, pages 147–168. Akademie-Verlag, Berlin, 2008.
- N. Calzolari, C. J. Fillmore, R. Grishman, N. Ide, A. Lenci, C. MacLeod, and A. Zampolli. Towards best practice for multiword expressions in computational lexicons. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, J. Odijk, and S. Piperidis, editors, *Proceedings of the Third*

International Conference on Language Resources and Evaluation (LREC-2002), Las Palmas, Canary Islands - Spain, 2002. European Language Resources Association (ELRA).

- K.-U. Carstensen, C. Ebert, C. Ebert, S. Jekat, R. Klabunde, and H. Langer. *Computerlinguistik und Sprachtechnologie: Eine Einführung.* Springer Spektrum, New York, 2010.
- P. Chilton. Analysing Political Discourse. Theory and Practice. Routledge, London, 2004.
- M. V. Dell'Anna. Lingua italiana e politica. Carocci Editore, Roma, 2010.
- S. Evert. Corpora and collocations. In A. Lüdeling and M. Kytö, editors, Corpus Linguistics. An International Handbook, pages 1212 – 1248. De Gruyter, Berlin, 2008.
- M. Foucault. *The archaeology of knowledge*. Tavistock Publications Limited, New York, 1972.
- A. Gardt. Diskursanalyse Aktueller theoretischer Ort und methodische Möglichkeiten. In I. H. Warnke, editor, *Diskursanalyse nach Foucault. Theorie* und Gegenstände, pages 27–52. de Gruyter, Berlin/New York, 2007.
- E. Goffman. On face-work: an analysis of ritual elements of social interaction. Pantheon, New York, 1989.
- R. Gould. The European paradox: Swiss discourses of identity between dependence and xenophobia. Language Ideologies, Policies and Practices. Language and the Future of Europe, pages 162 – 176, 2006.
- E. Guild. Security and Migration in the 21st Century. Polity Press, Cambridge, 2009.
- C. Hart. Critical Discourse Analysis and Cognitive Science. New Perspective on Immigration Discourse. Palgrave macmillan, Hampshire, 2010.
- A. Kilgarrif. Comparing corpora. International Journal of Corpus Linguistics, 6 (1):97–133, 2001.
- S. N. Kim. Statistical Modeling of Multiword Expressions. phdthesis, University of Melbourne, Melbourne, 2008.
- P. Koehn. Europarl: A parallel corpus for statistical machine translation, 2005.

- P. Koehn. Statistical Machine Translation. Cambridge University Press, 2010.
- P. Koehn. MOSES. User Manual and Code Guide. University of Edinburgh, 2016.
- L. Kueng. The impact of immigration on Swiss wages: A fixed effects two stage least squares analysis, 2005.
- D. M. Mayaffre and C. Poudat. Quantitative approaches to political discourse: Corpus linguistics and text statistics. In K. Fløttum, editor, *Speaking of Europe: Approaches to complexity in European political discourse*, pages 65–84. John Benjamins, 2013.
- T. McEnery and A. Hardie. The history of corpus linguisites. In A. Keith, editor, *The Oxford Handbook of the History of Linguistics*, pages 727–745. Oxford Handbook, Croydon, 2013.
- D. Orpin. Corpus linguistics and critical discourse analysis. examining the ideology of sleaze. *International Journal of Corpus Linguistics*, 10(1):37 61, 2006.
- E. Pianta, C. Girardi, and R. Zanoli. The Textpro tool suite. In '*LREC*', European Language Resources Association, Marrakech (Morocco), 2008.
- S. Pinker. The Blank Slate: The Modern Denial of Human Nature. Penguin, 2002.
- M. Popović, D. Stein, and H. Ney. Statistical Machine Translation of German Compound Words. In Advances in Natural Language Processing. 5th International Conference on NLP, FinTAL 2006 Turku, Finland, August 23-25, 2006 Proceedings, volume 4139 of Lecture Notes in Computer Science, pages 616 - 624. Springer, 2006.
- M. E. Saiz de Lobado and M. Bonomi. Metafora e vita quotidiana: l'immigrazione nella stampa italiana e spagnola. *Lingue Migranti e Nuovi Paesaggi*, pages 115–133, 2012.
- H. Schmid. Improvements in part-of-speech tagging with an application to german. In *Proceedings of the ACL SIGDAT-Workshop*, 1995.
- C. Schäffner and S. Bassnett. *Political Discourse, Media and Translation*. Cambridge Scholars Publishing, Cambridge, 2010.
- V. Seretan. Syntax-Based Collocation Extraction. Springer, Berlin, 2011.
- V. D. Teun A. Principles of critical discourse analysis. Discourse & Society, 4 (249):249–283, 1993.

- V. D. Teun A. What is political discourse analysis? *Political linguistics*, pages 11–52, 1997.
- L. Thomas, S. Kohls, and K. Krause. The labour market integration of immigrants and their children in Switzerland, 2012.
- E. Tognini-Bonelli. Corpus Linguistics at Work, volume 6 of Studies in Corpus linguistics. John Benjamins, Amsterdam, 2001.
- A. Triandafyllidou. The political discourse on immigration in southern Europe: a critical analysis. Journal of Community & Applied Social Psychologie, 10: 373–389, 2000.
- I. H. Warnke. Diskurslinguistik nach Foucault Dimensionen einer Sprachwissenschaft jenseits textueller Grenzen. In I. H. Warnke, editor, Diskursanalyse nach Foucault. Theorie und Gegenstände, pages 3–24. de Gruyter, Berlin/New York, 2007.
- M. Wengeler. Topos und Diskurs. Begründung einer argumentationsanalytischen Methode und ihre Anwendung auf den Migrationsdiskurs (1960-1985). Niemeyer, Tübingen, 2003a.
- M. Wengeler. Argumentationstopos als sprachwissenschaftlicher Gegenstand. Für eine Erweiterung linguistischer Methoden bei der Analyse öffentlicher Diskurse. In S. Geideck and W.-A. Liebert, editors, Sinnformeln. Linguistische und soziologische Analysen von Leitbildern, Metaphern und anderen kollektiven Orientierungsmustern, Linguistik – Impulse & Tendenzen, pages 59–82. de Gruyter, Berlin, New York, 2003b.

# Lebenslauf

#### Persönliche Angaben

Michela Rossi Ueberlandstrasse 391 8051 Zürich michela.rossi@uzh.ch

### Schulbildung

2010-2014	Bachelor-Studium Italianistik und Germanistik
	an der Universität Zürich
seit 2014	Master-Studium MLTA (Multilinguale Textanalyse) und Italianistik
	an der Universität Zürich

### Berufliche und nebenberufliche Tätigkeiten

09/2013 - 05/2014	Tutorat Cattedra di Letteratura Moderna
20/05/2014	Stewardess at AAIS
09/2014 - 12/2014	Tutorat Seminario di Letteratura Italiana Moderna
09/2015 - 12/2015	Tutorat Seminario di Letteratura Italiana Moderna
09/2015 - 12/2015	Tutorat Einführung in die Multilinguale Textanalyse
02/2016-06/2016	Tutorat Maschinelle Übersetzung und Parallele Korpora

# A Selection of Significant Extracted Elements

This appendix contains only a selection of our retrieved data, on the basis of the elements that we considered most relevant for the analysis. In the following tables we supply:

- 60 German most interesting keywords;
- 60 Italian most interesting keywords;
- 100 most interesting Intersecting keywords;
- 30 most interesting German 3-grams;
- 30 most interesting Italian 3-grams;
- 30 most interesting German 4-grams;
- 30 most interesting Italian 4-grams;
- 20 most interesting German 5-grams;
- 10 most interesting Italian 5-grams;

Every element is followed by Chi-squared value and number of occurrences in the primary political corpus.

### A.1 Keywords

German Keywords: Chi-squared value, Recurrences			
Zuwanderung (1514.9568950978341, 811) Asylmissbrauch (56.53208029345771, 30)			
Personenfreizügigkeit (652.9121947736968, 523)	kriminell (55.03792846158872, 144)		
Asylsuchende (553.9374465656072, 304)	Statistik (54.74941262277639, 73)		
Einwanderung (425.9267143992123, 226)	Tradition (53.24022031043919, 92)		
Masseneinwanderung (424.0418202311131, 225)	Sans-Papiers (48.7772967130434, 36)		
Flüchtling (415.2253727474484, 277)	Leib (48.7772967130434, 36)		
Person (400.43062735674096, 654)	Rückführung (46.668278742657, 44)		
Asylgesetz (363.5860294265295, 209)	Herkunftsland (46.41205835241409, 36)		
Migrant (343.268515927051, 193)	Missbrauch (43.853984627408806, 122)		
Arbeitskraft (248.82300392446214, 173)	Ausländerkriminalität (38.97680151853081, 43)		
Ausländer (222.7806959806082, 354)	Schlepper (36.53866449290738, 22)		
Steuerung (213.8245259130444, 143)	Attraktivität (32.222709804857836, 56)		
Kontingent (210.0946982946229, 138)	Asylproblem (32.034587545749254, 17)		
wir (175.23503996395434, 2404)	leben (32.01934335784759, 182)		
humanitär (138.03981282045976, 145)	Zielland (28.26577752432896, 15)		
steuern (132.18178944898008, 104)	Flüchtlingsstrom (27.951846520063526, 21)		
Schweiz (126.10469991803319, 3293)	Identität (27.216604851738715, 45)		
Grenze (106.41994411565341, 164)	Asylindustrie (26.38137601525299, 14)		
illegal (101.5826340324663, 114)	Asylmissbraucher (26.38137601525299, 14)		
Kriminalität (99.22571904411419, 82)	Illegale (25.359712720709165, 16)		
unser (92.48329582093393, 1583)	Sozialsystem (23.884659983622385, 24)		
Wohlstand (91.25026174161613, 147)	Migrationsstrom (24.496976840594, 13)		
Asylchaos (84.79890832202682, 45)	Wirtschaftsflüchtling (20.926973495010092, 12)		
eigenständig (81.08511813938317, 53)	Königsweg (20.83657587522593, 15)		
Schengen (79.43558619840228, 89)	Asylante (20.728185494509663, 11)		
renitent (77.81659082564498, 44)	Schönfärberei (20.728185494509663, 11)		
Verschärfung (73.49047561255665, 116)	Flüchtlingskontingent (19.049334919781476, 11)		
Beschleunigung (72.87411980363503, 50)	Scheinasylante (18.843793323075637, 10)		
Wirtschaftsmigrant (60.30096034862413, 32)	Zustrom (18.585991940762092, 18)		
EU-Bürger (59.901479978342955, 40)	Netto-Zuwanderung (16.959403486041257, 9)		

Table 33: German Keywords

Italian Keywords: Chi-squared value, Recurrences	
immigrazione (3220.274000373262, 1103)	vitto (107.08065784923765, 39)
immigrato (2200.9163650982487, 754)	albergo (97.96357676476742, 46)
clandestino (2107.0293655956903, 847)	hotel (96.99170546565584, 39)
migrare (1204.7980939261174, 416)	razzismo (91.9811354109627, 58)
profugo (1059.359768429323, 363)	terrorismo (85.8143571430281, 151)
accoglienza (1008.1222153477535, 413)	criminale (84.06238626059145, 128)
Nostrum (716.8958941238261, 276)	buonista (77.98111724193234, 33)
mare (611.3736800237322, 600)	sicurezza (72.46871509809228, 291)
invasione (533.8738511969416, 228)	tragedia (70.54893552473288, 76)
rifugiato (456.4182510724429, 158)	islamico (65.58907547307979, 104)
asilo (376.05214454629794, 187)	infiltrazione (59.603336851678236, 43)
sbarco (359.42037150144375, 138)	disperato (58.58218504349632, 38)
costa (290.0769549482535, 162)	irregolare (49.19420397766019, 27)
Europa (231.26201933155073, 807)	extracomunitario (47.92136737956092, 27)
frontiera (228.6491878033468, 135)	invadere (45.251580968040976, 30)
scafista (225.69139509798234, 91)	lsis (44.624818851885266, 85)
confine (217.7926357108157, 160)	stop (42.86534931671602, 94)
emergenza (213.67559350972073, 243)	controllo (42.00508301182934, 134)
Mediterraneo (212.40715299564198, 134)	scabbia (38.60452660712178, 14)
migratorio (199.49757133845117, 91)	rischio (33.45370440647454, 221)
flusso (186.17498151128297, 97)	lucrare (33.37008832089734, 23)
umanitario (177.4832764868113, 103)	lusso (32.44449565268775, 25)
business (156.8590423012141, 83)	esodo (29.147561262159684, 20)
morto (155.99692003501679, 143)	sedicente (25.48743845679054, 12)
espulsione (126.35430637054738, 70)	epidemia (24.028484622417178, 14)
migliaio (125.52298366813797, 189)	ebola (22.59611257443738, 12)
barcone (125.43778644444045, 57)	strage (20.380332503790505, 42)
rimpatrio (123.20815842784822, 46)	traversata (19.99102287261645, 10)
terrorista (116.00982171716754, 111)	carretta (19.99102287261645, 10)
alloggio (115.80895322200946, 59)	musulmano (18.395903875588843, 30)

Table 34: Italian Keywords

### A.2 Intersecting Keywords

German : Italian		
abkommen : stabilità	afrikanisch : africano	angst : paura
anlage : impianto	anwendung : applicazione	arbeitnehmer : lavoratore
arbeitslos : disoccupato	arbeitsmarkt : lavoro	asyl : asilo
asylsuchende : richiedente	aufenthalts- : soggiorno	aufgenommen : accogliere
aufnahme : accoglienza	aufnahme : ospitalità	aufnehmen : ospitare
ausgabe : spendere	ausländer : straniero	auswanderer : emigrato
auswanderung : migratorio	ausweisen : espellere	beschäftigung: occupazione
bevölkerung : popolazione	bilateral : bilaterale	brüssel : bruxelles
budget : bilancio	bundesgesetz : legge	demokratie : democrazia
dublin-abkommens : dublino	eu : ue	einwandern : immigrare
einwanderungsland : terra	fliehen : scappare	fliehen : fuggire
flucht : fuga	flüchtling : profugo	freizügigkeitsabkommen : circo- lazione
fremdenfeindlich : xenofobo	grenze : confine	grenze : frontiera
grenzraum : confine	grenzübertritt : traversata	hilfe : aiuto
hilfe : assistenza	humanitär : umanitario	illegale : clandestino
illegal : illegalmente	immigration : immigratorio	integration : integrazione
international : internazionale	investition : invetimento	jahr : anno
kriminell : criminale	küste : costa	land : paese
land : paese	landwirtschaft : agricoltura	leben : vita
medizinisch: sanitario	menschlich : umano	migrationsbereich : settore
migrationsstrom : migratorio	militärisch : militare	mittelmeer : mediterraneo
muslime : musulmano	nothilfe : emergenza	nigeria : nigeriano
person: persona	pferd : cavallo	produktion: produttiva
rechtschutz : tutela	reformprogramm : riforma	rückführung : respingimento
rückführung : rimpatrio	schlepper : scafista	schweizer : svizzero
schweiz : svizzera	stopp : stop	strom : flusso
syrer : siriano	syrien : siria	syrisch : siriano
terrorist : terrorista	trojanisch : troia	umsetzung : redistribuzione
umwelt : ambiente	ungebremst : incontrollato	unkontrolliert : incontrollato
unser : nostro	unterkunft : vitto	verkehr : circolazione
visum : soggiorno	volkswahl : elezione	volkswille : volontà
wirtschaftsflüchtling : rifugiato	zugewanderte : immigrato	zurückführen : respingimento
zuwanderer : immigrato	zuwanderung : migrazione	zwang : forzato

Table 35: Intersecting Keywords

### A.3 3-Grams, 4-Grams, 5-Grams

German 3-Grams: Chi-squared value, Recurrences	
[9 . Februar ]: 276.114608247, 180	
[in die Schweiz ]: 267.985928044, 216	
[mit der EU ]: 195.961160024, 265	
[Steuerung der Zuwanderung ]: 150.584708726, 80	
[Begrenzung der Zuwanderung ]: 120.601657742, 65	
[in unser Land ]: 96.4339216425, 59	
[aus der EU ]: 89.2676525785, 59	
[in der Schweiz ]: 68.6312775504, 717	
[Hilfe vor Ort ]: 57.5168486845, 38	
[Leib und Leben ]: 52.0104484361, 35	
[Steuerung und Begrenzung ]: 48.9383984183, 26	
[Neustrukturierung des Asylbereichs ]: 45.1738504422, 24	
[Attraktivität der Schweiz ]: 36.9122555154, 26	
[Beschleunigung der Verfahren ]: 36.1793501677, 21	
[Zahl der Asylgesuche ]: 35.7625211826, 19	
[in unserem Land ]: 29.9128912263, 102	
[Beschränkung der Zuwanderung ]: 26.3512500379, 14	
[Kontrolle der Zuwanderung ]: 26.3512500379, 14	
[Tradition der Schweiz ]: 25.4735807303, 25	
[Zuwanderung wieder eigenständig ]: 22.5867578522, 12	
[Senkung der Attraktivität ]: 22.5867578522, 12	
[Stopp der Zuwanderung ]: 22.5867578522, 12	
[Wohlstand der Schweiz ]: 21.9833222251, 17	
[Verschärfung des Asylgesetzes ]: 20.7045152461, 11	
[Schweiz als Zielland ]: 18.8222749646, 10	
[Grenzen wieder systematisch ]: 16.9400370077, 9	
[Italien und Griechenland ]: 11.2933370843, 6	
[der unkontrollierten Zuwanderung ]: 11.2933370843, 6	
[Einwanderung in unser ]: 11.2933370843, 6	
[die Rückgewinnung der ]: 11.2933370843, 6	

Table 36: German 3-Grams

Italian 3-grams : Chi-squared value, Recurrences
[reato di immigrazione ]: 149.458658791, 51
[sulle nostre coste ]: 136.809269407, 49
[migliaia di clandestini ]: 84.8973052605, 34
[vitto e alloggio ]: 60.0137867522, 22
[a casa loro ]: 56.3343631811, 20
[euro al giorno ]: 48.1712062741, 27
[migliaia di immigrati ]: 46.8880008998, 16
[reato di clandestinità ]: 44.3456891569, 26
[dei flussi migratori ]: 39.7192836467, 17
[morti in mare ]: 38.0964317798, 13
[sulle coste italiane ]: 37.6220982338, 15
[gli immigrati clandestini ]: 35.1659158117, 12
[bloccare le partenze ]: 35.1659158117, 12
[i nostri cittadini ]: 34.8553436497, 20
[a nostre spese ]: 32.2354033797, 11
[di infiltrazioni terroristiche ]: 29.3048944836, 10
[rischio di infiltrazioni ]: 27.0863848337, 10
[difesa dei confini ]: 26.3743891235, 9
[tragedia di Lampedusa ]: 23.4438872993, 8
[migliaia di morti ]: 22.1956337592, 9
[e propria invasione ]: 21.2530003905, 8
[criminale Mare Nostrum ]: 20.5133890111, 7
[a spese nostre ]: 19.3199160052, 8
[controlli alle frontiere ]: 19.3199160052, 8
[invasione di clandestini ]: 18.341837811, 7
[distruggere i barconi ]: 17.5828942589, 6
[parcheggio della disperazione ]: 17.5828942589, 6
[di carne umana ]: 17.5828942589, 6
[della nostra gente ]: 13.7554537777, 9
[politiche di invasione ]: 13.6154479792, 6

Table 37: Italian 3-Grams

German 4-grams : Chi-squared value, Recurrences
[Personenfreizügigkeit mit der EU ]: 76.7250877052, 46
[zur Steuerung der Zuwanderung ]: 56.4676576338, 30
[und Begrenzung der Zuwanderung ]: 52.8561821281, 29
[an Leib und Leben ]: 51.772170827, 31
[Zuwanderung in die Schweiz ]: 39.690463647, 22
[in die Schweiz kommen ]: 37.8102001051, 21
[der Personenfreizügigkeit mit der ]: 36.1794400195, 21
[die Personenfreizügigkeit mit der ]: 34.3055760683, 20
[Leib und Leben bedroht ]: 34.0502973231, 19
[Verträge mit der EU ]: 31.1430839221, 29
[die Hilfe vor Ort ]: 28.5915110761, 20
[Einwanderung in die Schweiz ]: 28.2335673007, 15
[humanitäre Tradition der Schweiz ]: 23.4767224614, 15
[generellen Stopp der Zuwanderung ]: 22.5868119984, 12
[die Kündigung der bilateralen ]: 22.5868119984, 12
[die Zuwanderung wieder eigenständig ]: 20.7045648802, 11
[der Einführung der Personenfreizügigkeit]: 18.8223200865, 10
[Grenzen wieder systematisch zu ]: 15.0578374728, 8
[die Einwanderung in die ]: 15.0578374728, 8
[ausser Kontrolle geratenen Zuwanderung ]: 13.1755996528, 7
[Arbeitskräfte aus dem Ausland ]: 12.4857525824, 9
[kriminelle und renitente Asylbewerber ]: 11.2933641573, 6
[Die heutige Masslosigkeit bei ]: 11.2933641573, 6
[der Neustrukturierung des Asylbereichs ]: 11.2933641573, 6
[Schweiz als Zielland für ]: 11.2933641573, 6
[die negativen Folgen der ]: 11.2933641573, 6
[Fachkräfte aus dem Ausland ]: 11.2933641573, 6
[gefährdet unsere Freiheit , ]: 11.2933641573, 6
[Volksinitiative "gegen Masseneinwanderung" . ]: 11.2933641573, 6
[Die Zahl der Asylgesuche ]: 11.2933641573, 6

Table 38: German 4-Grams

Italian 4-grams : Chi-squared value, Recurrences
[reato di immigrazione clandestina ]: 131.875202066, 45
[del reato di immigrazione ]: 67.4019864318, 23
[segretario federale della Lega ]: 59.5549836668, 73
[federale della Lega Nord ]: 58.1179066492, 71
[di migliaia di clandestini ]: 44.6269349796, 16
[a casa loro , ]: 23.4439570346, 8
[la casa di chi ]: 23.4439570346, 8
[vera e propria invasione ]: 21.2530657272, 8
[gestione dei flussi migratori ]: 21.2530657272, 8
[per bloccare le partenze ]: 20.5134500295, 7
[la nostra gente , ]: 15.4366760363, 6
[per i cittadini italiani ]: 15.4366760363, 6
[gli italiani in difficoltà ]: 14.652446627, 5
[politiche di invasione di ]: 14.652446627, 5
[la chiusura delle frontiere ]: 14.652446627, 5
[in mezzo al mare ]: 12.5406555128, 5
[clandestini nel nostro Paese ]: 11.7219502298, 4
[per difendere le nostre ]: 11.7219502298, 4
[un grande parcheggio della ]: 11.7219502298, 4
[La misura è colma ]: 11.7219502298, 4
[e militanti della cosiddetta ]: 11.7219502298, 4
[il leader dei Conservatori ]: 11.7219502298, 4
[confine con la Croazia ]: 11.7219502298, 4
[i propri confini e ]: 11.7219502298, 4
[di infiltrazioni terroristiche tra ]: 11.7219502298, 4
[rischio di infiltrazioni terroristiche ]: 11.7219502298, 4
[40 euro al giorno ]: 11.7219502298, 4
[il gioco dei terroristi ]: 11.7219502298, 4
[sbarchi sulle nostre coste ]: 11.7219502298, 4
[migliaia di vite umane ]: 11.7219502298, 4

Table 39: Italian 4-Grams

German t-grams : Chi-squared value, Recurrences	
[Steuerung und Begrenzung der Zuwanderung ]: 48.9386330552, 26	
[Wir machen uns stark für ]: 42.6621198355, 37	
[der Personenfreizügigkeit mit der EU ]: 34.305661418, 20	
[an Leib und Leben bedroht ]: 34.0503805585, 19	
[die Personenfreizügigkeit mit der EU ]: 32.4326355972, 19	
[eine harte , aber faire ]: 32.1708035473, 18	
[harte , aber faire Migrationspolitik ]: 28.233634984, 15	
[Leib und Leben bedroht sind ]: 24.4691201, 13	
[Steuerung und Kontrolle der Zuwanderung ]: 22.5868661449, 12	
[die Zuwanderung in die Schweiz ]: 22.5868661449, 12	
[weder einen generellen Stopp der ]: 20.7046145144, 11	
[einen generellen Stopp der Zuwanderung ]: 20.7046145144, 11	
[Kündigung der bilateralen Abkommen mit ]: 18.8223652085, 10	
[eigenständige Steuerung und Kontrolle der ]: 18.8223652085, 10	
[Attraktivität der Schweiz als Zielland ]: 15.0578735703, 8	
[Grenzen wieder systematisch zu kontrollieren ]: 15.0578735703, 8	
[für die Zukunft der Schweiz ]: 13.4162248328, 20	
[unseren Wohlstand in der Schweiz ]: 11.2933912304, 6	
[Landschaftsbild und letztlich unseren Wohlstand ]: 11.2933912304, 6	
[nicht an Leib und Leben ]: 11.2933912304, 6	

Table 40: German 5-Grams

Italian 5-grams : Chi-squared value, Recurrences
[il reato di immigrazione clandestina ]: 67.4021869241, 23
[centinaia di migliaia di clandestini ]: 32.2355951527, 11
[una vera e propria invasione ]: 15.4367240182, 6
[vitto, alloggio, sigarette]: 14.6524902118, 5
[per bloccare le partenze , ]: 14.6524902118, 5
[per difendere le nostre frontiere ]: 11.7219850975, 4
[esponenti e militanti della cosiddetta ]: 11.7219850975, 4
[questo governo e questa maggioranza ]: 11.7219850975, 4
[cancellazione del reato di immigrazione ]: 11.7219850975, 4
[la salute dei nostri cittadini ]: 11.7219850975, 4

Table 41: Italian 5-Grams

# **B** List of Corpora and Tools

This appendix contains a list of corpora and tools that we used for this project.

- Corpora
  - bulletin4corpus (http://kitt.cl.uzh.ch/kitt/b4c/en)
  - Eroparl (http://www.statmt.org/europarl/)
  - German Political Corpus
  - Italian Political Corpus
- Tools
  - import.io (https://www.import.io/)
  - jWordSplitter (http://www.danielnaber.de/jwordsplitter/index\_en. html)
  - Moses (http://www.statmt.org/moses)
  - TextPro (http://hlt-services2.fbk.eu/textpro/)
  - TreeTagger (http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/)



Philosophische Fakultät Studiendekanat

Universität Zürich Studiendekanat Rämistr. 69 CH-8001 Zürich www.phil.uzh.ch

#### Selbstständigkeitserklärung

Hiermit erkläre ich, dass die Masterarbeit von mir selbst und ohne unerlaubte Beihilfe verfasst worden ist und ich die Grundsätze wissenschaftlicher Redlichkeit einhalte (vgl. dazu: http://www.lehre.uzh.ch/plagiate/20110314\_LK\_Merkblatt Plagiat.pdf).

Zunch 08/10/16 Unchela Biss

Ort und Datum

Seite 1/1 11.12.2015