



# I-CLAIM

Improving the Living  
and Labour Conditions  
of Irregularised Migrant  
Households in Europe

Methodological Note:

## **Corpus-based discourse analysis of migration-related discourses in media, politics and civil society**

---

Markus Rheindorf, Bastian Vollmer

February 2025



Funded by  
the European Union



UK Research  
and Innovation

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or UK Research and Innovation. Neither the European Union nor the granting authority can be held responsible for them.

|                      |                                   |
|----------------------|-----------------------------------|
| Authors:             | Markus Rheindorf, Bastian Vollmer |
| Final Reviewers:     | -                                 |
| Dissemination Level: | -                                 |

#### History of changes

| Version   | Date     | Change |
|-----------|----------|--------|
| Version 2 | 1-2-2025 |        |

## Table of Contents

|   |           |
|---|-----------|
| <b>1. Introduction .....</b>                              | <b>4</b>  |
| <b>2. Study Design .....</b>                              | <b>5</b>  |
| 2.1. Data Collection.....                                 | 5         |
| 2.2. Quantitative analysis.....                           | 6         |
| 2.2.1. Compiling the corpus and sub-corpora .....         | 6         |
| 2.2.2. Frequency analysis.....                            | 6         |
| 2.2.3. Semantic grouping .....                            | 7         |
| 2.2.4. Collocation analysis .....                         | 8         |
| 2.2.5. Comparing corpora and keyness analysis .....       | 9         |
| 2.3. Downsampling for qualitative narrative analysis..... | 10        |
| 2.3.1. Building proto-narratives.....                     | 10        |
| 2.3.2. Downsampling.....                                  | 11        |
| 2.4. Qualitative analysis of narratives .....             | 11        |
| <b>3. References .....</b>                                | <b>17</b> |

## 1. Introduction

This methodological note lays out some fundamental concepts, definitions and procedures used in the corpus-based discourse analysis conducted in Work Package 4.

I-CLAIM conceptualises irregularity as produced at the nexus of nested legal systems, political and public discourses on irregularity. Work Package 4 focuses on the discursive dimension of the irregularity assemblage and, in the survey part, the perception of discourse. We refer to the ‘discursive construction’ and to the ‘discursive representation’ of social realities, such as irregularised migrants, migration and irregularity as such. WP4 focuses on discourse in media, politics and civil society as key areas of the discursive construction of irregularity.

### Discourse

From the perspective of WP4, the following aspects of the irregularity assemblage are crucial: Discursive constructions of irregularity and associated entities, actions and people, in particular through linguistic interactions manifest as ‘texts’, both reflect and constitute a significant part of the social reality of irregularity. We regard discourse as a polyphonous, multi-voiced field of social interaction and thus also of the contestation of social meanings. Discourse, in other words, has ontological power (Vollmer, 2021).

In short, even if there is a dominant or hegemonic meaning or narrative of irregularity, there will be alternative or counter-hegemonic voices, positions and narratives. Through a large sample of textual data, we aim to include a wide range of heterogeneous actors’ contribution to the discourse about irregularity. By including politics, media and civil society, a comparative perspective is facilitated. Regarding intersectionality, we aim to explore the textual with respect to interdiscursive links between gender, race, and class. However, interdiscursivity may also reveal very specific combinations characterising a specific country’s irregularity assemblage, e.g., regarding the link between migration and labour markets or between migration and national identity.

### Narrative

We define ‘narrative’ as a textual sequence of meaningful events that orders reality in a comprehensible way. They inevitably leave out certain elements (e.g. plots and figures) in their telling and focus on particular other elements by which they evaluate or judge aspects of reality. This ‘ordering’ or ‘transformation’ we refer to as ‘emplotment’. This kind of emplotment has been studied as a way of encoding ideology or *Weltanschauung* into even the simplest story. Key elements of narrative in this view, as in many others, are characters (often but not exclusively ‘hero’, ‘villain’, ‘helper’ and ‘bystander’), setting or circumstance (including places, time periods), plot (the sequence of actions or events), means (tools, devices that help or hinder characters) and moral (evaluation or judgment). This basic pattern is not always fully set out and may also be ambiguous and therefore open to differing perceptions and varying interpretations.

### Objectives

The corpus analysis, using statistical methods developed in Corpus Linguistics, allows us to understand the lexical patterns of large sets of linguistic data, thus approximating discourses more comprehensively than in-depth small-scale qualitative analysis. The lexical patterns we aim to identify concern the people, entities and processes involved in the irregularity assemblage in specific country contexts – see below for details on methodology.

Based on the statistical results of the corpus analysis, we down-sample to much smaller samples of texts to analyse qualitatively: the lexically typical texts (averages) and the very unusual (outliers). Here, the focus is on understanding the narratives around irregularity and the representation of the social actors, actions and entities involved – using a combination of narrative analysis and Social Actors Analysis, see below.

On both the quantitative and the qualitative level, we compare and contrast the discursive construction of irregularity in the media, politics and civil society to identify the specificity of discourses and narratives in each domain, in particular to find alternative representations and narratives about irregularised migrants/migration in civil society (Baker et al. 2008).

## 2. Study Design

### 2.1. Data Collection

Each participating country team compiled a separate corpus for the following 3 domains: media, politics and civil society. The time period considered was roughly the last 5 years, i.e., 2019-2023. Data collection was designed to be extensive for the quantitative approach, so as to allow us to catch the patterns that are too large to see with the naked eye. Texts were saved in .txt format for analysis, and we used file names that contain key meta information, such as source name and date, as in “Bild-10-01-2023”. All texts for one of the 3 corpora were then saved/stored in the same dedicated folder.

The media data was compiled to reflect the media landscape of each country in terms of circulation, quality of publication (tabloid – quality), political leaning (progressive – conservative) and regional reach (regional – national). No media data was collected for the EU level. Given the varying sample sizes, comparisons between groupings (quality vs tabloid or progressive vs conservative) uses normalised statistics. The aim was to choose at least 2 national newspapers with wide reach, where relevant complement this also with weekly/monthly magazines or online news outlets. We aimed for a mix of tabloid and broadsheet/quality, left- and right-leaning, where applicable. From each pre-selected media source, we collected texts addressing salient issues in the respective country. Since we could not standardise search terms or topics across all countries – beyond ‘irregular migration’ – specific search terms were chosen by each team. Considering the time period covered, however, we expected the pandemic and Russia’s war on Ukraine to be significant internationally.

For the politics corpus, each team collected parliamentary debates (or equivalents), press releases of governments and government agencies, as well as of opposition parties, that address salient issues in the respective country. Topic-wise, the same applies as for the media. We aimed to make sure that salient topics with a bearing on irregular migration in the respective country are included, but expected the pandemic and the war in Ukraine to be relevant. This corpus was designed to in particular include debate over policies and laws related to irregularised migrants/migration. With parliamentary debates, it was not be possible to neatly separate speakers, so entire sessions were saved in one file. At the EU level, the European Commission and the European Parliament were represented in the politics corpus.

For the domain of civil society, finding and selecting of sources was expected to be most challenging. We aimed to include NGOs and other civil society organisations that work with and speak about irregular migrants, including but not limited to those organisations in our respective stakeholder groups. Websites, newsletters, press releases, social media posts etc. were included. To ensure that the search terms we used

to find texts for the quantitative analysis do not exclude some of the alternative/counter-hegemonic narratives used by some of the actors to advocate for migrants with precarious status, we suggested two strategies: First, draw on stakeholder partners to identify alternative terminology that, for example, avoids hegemonic terms like ‘irregular’ or ‘illegal’. Second, regarding the civil society corpus, call on stakeholders to send in/suggest texts as well; these will have to be vetted before inclusion in the corpus.

Regarding the weight given to individual sources (e.g., newspapers) in terms of how many articles or words we include from each, we aimed for rough parity between sources on the same level. For example, between all the parties represented in parliament; or between all national newspapers. However, with smaller news outlets or monthly magazines, such numerical parity might well be impossible and adhering to it strictly would defeat the purpose of our data collection.

## 2.2. Quantitative analysis

For the corpus linguistics analysis we used the software AntConc, as it is reliable, free of charge. The following steps outline the key stages of CL analyses. Full documentation can be found here: <https://antconc-manual.readthedocs.io/en/latest/>

### 2.2.1. *Compiling the corpus and sub-corpora*

In the Corpus Manager menu, each team uploaded all texts or select the folder containing all files, then named and saved the corpus for easy access later. We will call this the main corpus. This submenu also contains some key settings and parameters such as the token definition (what will count as a unit in the analysis). We will keep the default settings here.

To compare individual sources (e.g. politics vs media, or more specifically tabloid vs broadsheet newspapers) and to discover the unique characteristics of any individual source or group of sources, we created corpora (subcorpora in our case, because they consist of elements from the main corpus) that consist of only the texts from those sources—in addition to the main corpus. We created subcorpora for politics, media and civil society.

### 2.2.2. *Frequency analysis*

The simplest and usually first step in CL analyses is to generate a frequency list. This computes the frequency of each identifiable token in the corpus (i.e. discrete units separated with a space, including words, numbers, etc.) and lists these tokens by frequency. Note that different spellings and forms of the same lexeme will initially be counted as different tokens. “Two” is counted as different from “2”; in English, “colour” is counted as different from “color”, and “travel” is counted as different from “travelling”; and so forth. At the same time, identical forms with different meanings are counted as the same, e.g., the noun “attack” and the verb “attack”. Figurative, metaphorical or ironic usage of words is not recognised by the compiler either. These are some of the key limitations of CL and need to be kept in mind in the later stages of analysis.

Typically, frequency lists contain among their top ranks many function words like articles, auxiliary verbs, conjunctions and pronouns. While these may be interesting in some situations, they are usually discarded / ignored in topic-focused analyses. Indeed, we suggest you focus on content words like ‘migrant’ or ‘refugee’. A practical approach is to save/export the frequency list and edit it in Excel by removing said function words, ‘pruning’ the list. Copying the result into another a fresh sheet, you may then want to combine tokens that are spelling variants or forms of the same lexeme, summing up their separate frequencies, and then resorting the list by the combined frequencies.

Frequency lists of large corpora run to many thousands of entries, so that you may not want to continue this process for the entire list but stop at words with a low frequency (usually a low normalised frequency like 1 per 10.000 or 1.000 words). This will yield a good overview of the dominant topics/words. However, since we are also interested in marginal(ised) topics and near-silences, you should at least parse the first 1000 items on the list, ignoring function words like articles and prepositions, and note relevant tokens with low frequencies in that respect. Later, comparisons between the subcorpora can also help identify silences in one source/group compared to others, e.g., civil society vs. politics.

While each team explored the frequency lists without an exclusive focus, to discover dominant topics, the following questions were suggested to guide this exploration:

- Do the sources refer to specific groups of migrants?
- Which terms are used to refer to irregularised migrants?
- What political, governmental, welfare etc. institutions or organisations are referred to?
- Which actions/activities do the sources refer to?
- Which qualities/characteristics do the sources refer to?
- Which locations, places or sites do the sources refer to?

### 2.2.3. *Semantic grouping*

If a lexical field (words referring to the same or similar things) is highly dispersed, the individual frequency of each token will be low – perhaps too low to be noticed. Semantic grouping seeks to prevent this problem by clustering terms into semantically close/related groups, e.g., all terms referring to a government, to the welfare state or to migrants. Then, the individual frequencies for all terms in a group are summed up to see the actual presence of the entity, with all its ‘names’ (Orpin 2005; McEnery/Gabrielatos 2006; Baker/McEnery 2005).

After discussing our initial findings, the work package lead-team suggested the following semantic groups be used by all teams, in addition to any specific sub-groups or additional semantic groups found by individual teams. We thus have a comparable core of semantic group across all data sets.

#### Recommended semantic groups

**Labour** (including economy or economy as subcategory)

**Migration and asylum** (system, process), including the following potential subgroups:

- Borders
- Arrival
- Return
- Asylum

**Status** (the quasi-legal and legal migration status of individuals and groups, including ‘migrant’, ‘those arriving’)

**Geography** (places, place names, generic references like city or country, national, regional and local)

**Rights** (human rights, specific legal rights)

**Welfare** (actual welfare in practice, not rights underlying the welfare system)

- Housing
- Education
- Health (all health and healthcare related terms)

**Subject** (the subjective experience and lived reality of individuals and groups)

- Social categories/identities: demographics, incl. age, race, religion, gender, sexual orientation
- Family and household: parent, father, mother, child, home, friend etc.
- Subjective experience: life, living, eat, sleep, feeling, thinking, emotions etc.

**Politics and governance** (international, national and local)

- Policy/rules and legislation
- The state and its institutions (gov., ministries, other state authorities, elections, parties, politicians)
- NGOs, INGOs, IOs,
- Law and order (police, policing, protests etc.)

**Crime** (criminal actions, crimes, victims, perpetrators)

**Quantities** (including numbers and noncountable nouns: hundreds, thousands, masses, flood, wave, increase, decrease)

**Crisis** (crisis, catastrophe, threat, burden, collapse etc.)

All teams created sum frequencies for their semantic groups to compare prevalence of topics, i.e., the semantic preoccupations of the respective corpora.

#### 2.2.4. *Collocation analysis*

One of the key advantages of CL is that it allows us to discover patterns of language use across large data sets that are 'invisible' to the naked eye – not simply because of the large number of texts but because these patterns exceed our attention span and capacity to separate coincidental from significant co-occurrence. This is very evident in the case of collocations, the neighbours or co-text of specific words. In other words, collocational analysis shows us which words are combined in a given corpus; for instance, which adjectives/qualities are commonly attributed to 'migrant'? Which actions/verbs are commonly linked to 'migrant'? (Orpin 2005; McEnery/Gabrielatos 2006; Baker/McEnery 2005).

For the most salient words (there should be a few in each semantic group), we identified collocations and, where relevant to I-CLAIM's research focus, built semantic groups for those collocations (summarising not frequency but likelihood). These salient words related to irregularity, migration, etc., and do not include unrelated words that, for some reason, are very frequent (in the German politics corpus, 'applause' is very common but unrelated, so it was not included in this step).

However, the statistical procedures underlying collocational analysis go beyond simply counting the frequency of such word combinations. Using one of several tests, collocational analysis calculates significant



combinations beyond the mere likelihood or coincidence of two words ‘happening to be used next to each other’, thus giving us a reliable picture of significant combination.

Collocations are thus patterns that are habitual or characteristic of a particular discourse, medium or domain of language use (such as ‘politics’). These combinations have been shown to have a priming effect (lexical priming) that will lead people to associate a word with its habitual combinations even if those combinations are not present in a specific text/sentence: if ‘criminal’ or ‘violent’ become established as habitual combinations of ‘migrant’, these qualities will tend to be associated with ‘migrant’ even in their absence. This effect has been described as the ‘aura’ of a specific word.

Collocations are always computed for a search term, but that search term can be single word or a string of tokens. The search window around the search term determines the range (in number of tokens) which is used to compute collocations. The results are shown by strength of collocational pattern, which related to frequency of occurrence but not identical to it.

Collocations are especially interesting around words denoting migration and migrants (including loose synonyms, because these give a picture of typical associations/stereotypes and narratives. In particular, we want to consider verbs, adjectives and nouns among these collocations, i.e., the actions, qualities and entities commonly associated with migrants and migration.

Within semantic groups, a comparison between individual words was optionally conducted. For example, ‘Covid’ might be the most frequent word within the ‘health/healthcare’ semantic group; ‘integration’ might be the most frequent within ‘obligations’; ‘rape’ might be the most frequent within the group ‘crime’ etc.

For all frequency values (=words and semantic groups), we also calculated the normalised value per 1000 words in the respective corpus. Thus, if ‘migrant’ occurs 5,500 times in 10,000,000 tokens, the normalised frequency would be  $5,500 / 10,000,000 * 1000$  (raw frequency divided by corpus size multiplied by 1000 as our reference value).

### 2.2.5. *Comparing corpora and keyness analysis*

In statistically comparing two corpora/data sets on a lexical basis, AntConc compares the normalised frequencies of each token found in either of the two data sets. Note that this is always a two-point comparison, e.g., between two subcorpora or between the main corpus and one subcorpus. For our purposes, both approaches can be used. The keyness comparison between the politics and media subcorpora, for example, will reveal what makes either different from the other lexically; while the comparison between the civil society corpus to the main corpus will reveal what is unique about the civil society corpus among all the data we collected (i.e. different from politics and media together). We call the corpus we focus on the primary corpus and the corpus it is compared to the reference corpus in that situation.

Keyness can be a positive or negative value, but it is relative and therefore does NOT follow a standard scale (of 1-100 for example). Positive keyness values, e.g., 428, indicate how a given word is more characteristic of the primary corpus relative to its reference corpus. Negative keyness, e.g., -264, indicates how uncharacteristic the word is of the primary corpus relative to the reference corpus. Taken together, the words with high keyness then characterise the distinguishing lexis and, by implication, topics, of the primary corpus relative to the reference corpus. (Orpin 2005; McEnery/Gabrielatos 2006; Baker/McEnery 2005)

All teams carried out keyness comparison for their corpora, i.e., 3 two-way comparisons for a total of 6 comparisons. Beyond keyness, we also compared (manually) the following: frequency list-based semantic groups (their sum frequencies and the internal composition of groups); the collocations of our focus words (e.g. migration, migrants); and the collocations-based semantic groups of those words. This helps identify typical and diverging narratives.

## 2.3. Downsampling for qualitative narrative analysis

### 2.3.1. Building proto-narratives

Based on the quantitative results indicating a link to ‘narratives’, see above, and centred on our focus words denoting migration and migrants, we constructed proto-narratives to be explored in detail. These were constructed around base words, i.e., salient terms in the respective data sets that relate to I-CLAIM’s research focus, such as ‘migrant’, ‘migration’ and ‘labour’. Proto-narratives are thus analytical abstractions based on the understanding of narrative as typically comprising (1) characters/actors, (2) actions/activities, (3) settings or circumstances, (4) means, tools and objects engaged with, (5) a moral or evaluation (Wodak/Rheindorf 2017).

For example, a strong collocation (high likelihood) of ‘migrants’ with the verbs ‘live’, ‘resist’ and ‘deported’ suggests 3 main proto-narratives around migrants living, resisting and (being) deported; collocations with the words ‘border’, ‘underground’ and ‘cities’ suggest that these might be the main settings/circumstances etc. Narratives can centre not only on human beings (migrant) but also on abstract concepts (migration) or organisations (the government), using rhetorical devices such as pars-pro-totem or metonymy (as in ‘Berlin has decided’ or ‘Migration has changed the face of Germany forever’).

In some narratives, ‘migrants’ are grammatically active and in others passive, i.e., they act or are acted upon. These steps can also be used to find different proto-narratives for different groups of migrants, distinguished by origin, gender, race, religion etc. (e.g. Afghan vs Ukrainian), if such groups appear salient in your country-specific data. The following examples illustrate the composition of such proto-narratives.

|                         |                              |                    |                                    |                                     |
|-------------------------|------------------------------|--------------------|------------------------------------|-------------------------------------|
| An increasing number of | irregular migrants           | are coming to      | Germany                            | via the Balkans route.              |
| <i>Attribute</i>        | <i>character</i>             | <i>plot/action</i> | <i>setting</i>                     | <i>means/setting</i>                |
| The Germany economy     | needs and has benefited from | migrants           | and is lobbying for regularisation | in a self-interested and amoral way |
| <i>character</i>        | <i>plot/action</i>           | <i>plot/action</i> | <i>plot/action</i>                 | <i>moral/evaluation</i>             |

The steps above were carried out for each corpus, as the narratives might be very different between corpora. At the end of these steps, a small number (3, no more than 5) of salient proto-narratives for each of your country-specific focus words was formulated for each corpus.

### 2.3.2. *Downsampling*

Using the KWIC concordance viewer in AntConc, each team retrieved a small sample of concrete texts that feature realisations of the proto-narratives. Such realisations are, by definition, more fully fleshed-out variations of the abstract proto-narratives previously constructed (Rheindorf 2019a).

## 2.4. Qualitative analysis of narratives

Based on the small sample thus obtained, we move beyond the proto-narratives we constructed before, identifying the fully realised narratives in the material. In the following qualitative analysis, the narrative elements (characters, settings, plot/actions, means and moral) are identified in each of these downsampled text, but only for the narratives centred on the respective base words (e.g., ‘migrants’ or ‘migration’). Additional narratives that also occur in these texts were included only if they were connected with or had a clear bearing on the narrative the text had been selected for. Particular attention was paid to the representation and narrative framing of irregularised migration and migrants, their labour conditions and household/family lives. The preliminary result of this qualitative approach is an analytical characterisation of main narratives for each corpus, with reference to the specific texts we found them in and/or their authors (e.g., newspaper, genre, date of publication, politician or institution etc.).

Even with this basic definition, much of the data collected cannot be classified as narratives. Instead, they might for example be descriptions of states or conditions. The method of Social Actors Analysis (SAA) provides both a more fine-grained analysis of narratives (and their components) as well as insight into non-narrative forms that relate to the same characters, settings, plots, means or morals. SAA was suggested as a shared frame of reference and terminology for the qualitative analysis.

While other forms of SAA are used in sociology, we refer here to the socio-semiotic approach initially developed by van Leeuwen (1996, 1999, 2008) as a way of systematically analysing the representation of social actors in texts, linguistic and otherwise. Social actors do not necessarily map onto grammatical actors; for instance, impersonalized actors can be non-human entities that are still represented as engaged in particular actions, be it as active actor or as passive goal; while passive constructions can background actors, and agent deletion (in English passive) can suppress the actor entirely. Similarly, actions can be linguistically represented as states (through nominalisation), etc. For our purposes, this is how we propose to study the employment of narratives and other forms.

SAA provides us with a systematic way of doing so and the corresponding terminology to identify and name these transformations. The essential tool for this are a number of ‘system networks’ that basically work like decision trees. The figure below maps the possible transformations in the representation of entities (people, organisations etc.), with square brackets representing either-or choices and round brackets complementary choices in the representation. The numbers indicate the recommended sequence of identifying these transformations.

To illustrate, consider the headline ‘Eleven dead in riots’ referring to an event in Apartheid-era South Africa. During a large-scale protest, police opened fire on the protesters, killing 11 and injuring dozens more, all of whom were black. The sentence in question presents a specific version of these events: (1) The people killed are included, (2) those doing the killing are excluded completely, i.e., suppressed. A hypothetical alternative sentence might have simply backgrounded them by, e.g., referring to police cars. (3) The included are neither active or passive, as the sentence ellipsis lacks a verb. (4) The riots are presented as a circumstance, a setting

without anyone participating or owning/possessing it. (6) The identified are presented simply by the number '11', which is impersonalisation – abstraction (22). An alternative sentence might have identified them as 'blacks', 'protesters', 'South Africans', 'men and women' etc., constituting personalisation (10) – determination (13), followed by various forms of categorization or nomination (13).

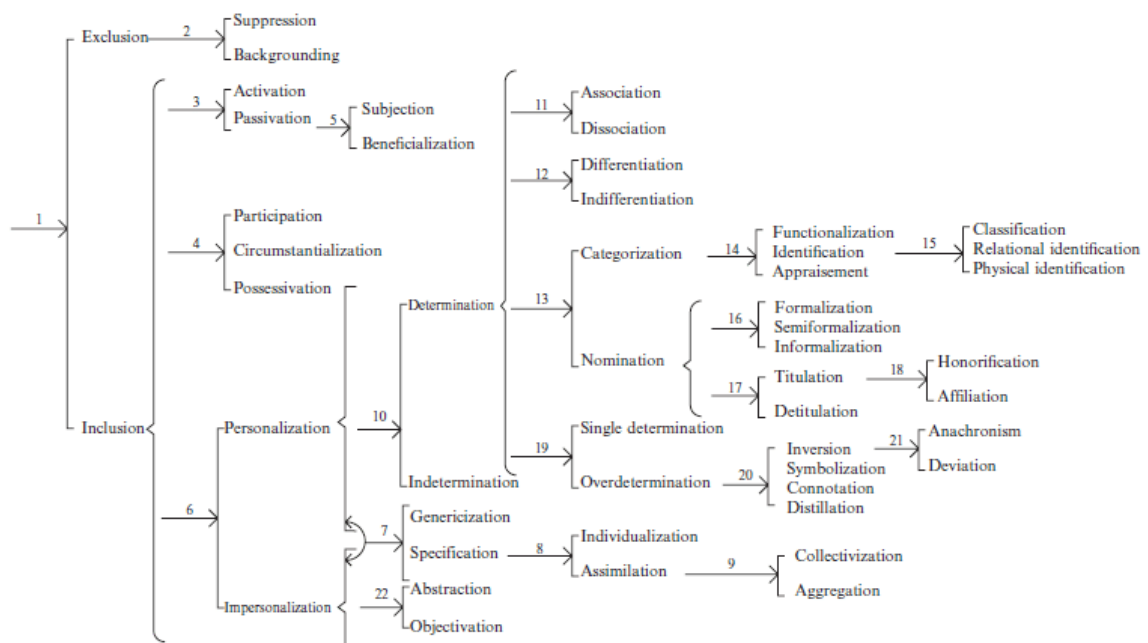


Fig 1: Linguistic options in the representation of social actors

Each of the categories is defined in distinction to the other categories of this network. In the following we give a brief definition, but please refer to the literature for more in-depth discussion and examples.

1. A social actor is either included in or excluded from the linguistic representation of an event or action.
2. Exclusion: a social actor/entity is completely deleted (suppression) or lexico-grammatically pushed into the background (backgrounded).  
  
Included actors are subject to (3) activation/passivation, (4) participation/circumstantialization/possessivation, and (5) personalization/impersonalisation – all at the same time.
3. Inclusion: a social actor is considered activated when presented as agent (semantic role as agent) in a sentence, even if that sentence is in passive voice (grammatical role); a social actor is considered passivated when presented as patient/recipient (semantic role), even if they are the grammatical subject of that sentence (grammatical role).
4. Inclusion: a social actor is presented through participation when presented as participating in an event/action rather than originating it or receiving it; circumstantialization occurs, when a social actor is presented as a bystander, e.g., witness, to an action or event; possessivation occurs, when a social actor is presenting as 'owning' or controlling an action, entity or event.

5. Within passivation, we distinguish between subjection and beneficialisation; broadly speaking, the former covers actions with negative consequences, the latter positive ones that the social actor receives (is semantic patient to).
6. Personalization presents social actors as human beings, as realised by personal or possessive pronouns, proper names, or nouns (and sometimes adjectives as, for example, in “maternal care”) whose meaning includes the feature “human.” But social actors can also be impersonalized, represented by other means, for instance, by abstract nouns or by concrete nouns whose meanings do not include the semantic feature “human.”
7. The choice between generic and specific reference is another important factor in the representation of social actors; they can be represented as classes, or as specific, identifiable individuals. Generic reference can take the form of plural nouns without article (migrants), singular nouns with definite or indefinite article referring to a type (the/a migrant as a class) or mass nouns (police).
8. In the case of specification, social actors may be individualised or assimilated. Social actors can be referred to as individuals, in which case I shall speak of individualization, or as groups, in which case I shall speak of assimilation.
9. SAA distinguishes two major kinds of assimilation: aggregation and collectivization. The former quantifies groups of participants, treating them as statistics, the latter does not.
10. Determination/Indetermination: Indetermination occurs when social actors are represented as unspecified, “anonymous” individuals or groups, determination when their identity is, one way or another, specified. Indetermination is typically realized by indefinite pronouns (“somebody,” “someone,” “some,” “some people”) used in nominal function.:
11. Another way in which social actors can be represented as groups is by association. Association refers to groups temporarily formed by the text to include some social actors and/or groups of social actors (either generically or specifically referred to). The most common realization of association is parataxis: “children, families and women”.
12. Differentiation/Indifferentiation: Differentiation explicitly differentiates an individual social actor or group of social actors from a similar actor or group, creating the difference between the “self” and the “other,” or between “us” and “them,” as with “others”.
13. Social actors can be represented either in terms of their unique identity, by being nominated, or in terms of identities and functions they share with others (categorization), and it is, again, always of interest to investigate which social actors are, in a given discourse, categorized and which nominated. In stories, for instance, nameless characters fulfill only passing, functional roles and do not become points of identification for the reader or listener. In press “stories,” something similar occurs. A newspaper may, for instance, nominate only a high-status person, a government minister. The press (and not only the press) also tends to nominate men and women in different ways, for instance by referring to marital status only in the case of women.

Nomination is typically realized by proper nouns, which can be formal (surname only, with or without honorifics), semiformal (given name and surname, as with “Boris Johnson”), or informal (given name only, as with “Boris”).

Items other than proper names may be used for nomination, especially when, in a given context, only one social actor occupies a certain rank or fulfills a certain function. Nominations of this kind in fact blur the dividing line between nomination and categorization. They are common in stories for young children, with characters referred to as “the Little Boy,” “the Giant,” “the Rabbit,” etc., even in vocatives:

14. There are two key types of categorization: functionalization and identification. Functionalization occurs when social actors are referred to in terms of an activity, in terms of something they do, for instance, an occupation or role. It is typically realized in one of the following ways: first, by a noun, formed from a verb, in English through suffixes such as -er, -ant, -ent, -ian, -ee, e.g., “interviewer,” “celebrant,” “correspondent,” “guardian,” “payee”; second, by a noun which denotes a place or tool closely associated with an activity, in English through suffixes such as -ist, -eer, e.g., “pianist,” “mountaineer”; third, by the compounding of nouns denoting places or tools closely associated with an activity (“villager”) and highly generalized categorizations, such as “man,” “woman,” “person,” “people”, and functionalizations such as “cameraman,” “chairperson.”

Identification occurs when social actors are defined, not in terms of what they do, but in terms of what they, more or less permanently, or unavoidably, are. I have distinguished three types: classification, relational identification, and physical identification.

In the case of classification, social actors are referred to in terms of the major categories by means of which a given society or institution differentiates between classes of people. In the West, these now include age, gender, provenance, class, wealth, race, ethnicity, religion, sexual orientation, and so on. But classification categories are historically and culturally variable. What in one period or culture is represented as “doing,” as a more or less impermanent role, may in another be represented as “being,” as a more or less fixed identity.

Nominations may be titulated, either in the form of honorification, the addition of standard titles, ranks, etc., as with “Dr.”, or in the form of affiliations, the addition of personal or kinship relation terms, as with “Auntie Barbara”.

15. In the case of classification, social actors are referred to in terms of the major categories by means of which a given society or institution differentiates between classes of people. In the West, these now include age, gender, provenance, class, wealth, race, ethnicity, religion, sexual orientation, and so on. But classification categories are historically and culturally variable.

Relational identification represents social actors in terms of their personal, kinship, or work relations to each other, and it is realized by a closed set of nouns denoting such relations: “friend,” “aunt,” “colleague,” etc. Typically, they are possessivated, either by means of a possessive pronoun (“her friend”) or by means of a genitive (“the child’s mother”) or postmodifying phrase with of (“a mother of five”).

Physical identification represents social actors in terms of physical characteristics which uniquely identify them in a given context. It can be realized by nouns denoting physical characteristics (“blonde,”

“redhead,” “cripple,” and so on) or by adjectives (“bearded,” “tall”) or by prepositional phrases with or without postmodifying highly generalized classifications such as “man” or “woman.”

16. Formalization/Semi-formalization/Informalization refers to the degree of formality conferred to the representation of a social actor.
17. Closely connected to the above, including or omitting a title in the representation of a social actor can be important.
18. This may use honorifics or happened through affiliation with a prestigious institution etc.
19. Not relevant to our analysis.
20. Not relevant.
21. Not relevant.
22. There are two types of impersonalization: abstraction and objectivation. Abstraction occurs when social actors are represented by means of a quality assigned to them by and in the representation. One example is the way in which “poor, black, unskilled, Muslim, or illegal” immigrants are referred to by means of terms like “problem”, “crisis” or “wave”. They are assigned the quality of being problematic, a threat or unstoppable force, and this quality is then used to denote them.

Objectivation occurs when social actors are represented by means of reference to a place or thing closely associated either with their person or with the action in which they are represented as being engaged. In other words, objectivation is realized by metonymical reference. A number of types of objectivation are particularly common: spatialization, utterance autonomization, instrumentalization, and somatization.

Spatialization is a form of objectivation in which social actors are represented by means of reference to a place with which they are, in the given context, closely associated. This happens, for instance, when “Germans” are substituted by “Germany.”

Utterance autonomization is a form of objectivation in which social actors are represented by means of reference to their utterances. This is the case, for instance, when “the report” and “surveys” is said to “report” something. This lends a kind of impersonal authority to the utterances, it is often used in connection with the utterances of high-status and “official” spokespeople.

Instrumentalization is a form of objectivation in which social actors are represented by means of reference to the instrument with which they carry out the action in which they are represented as being engaged.

Somatization, finally, is a form of objectivation in which social actors are represented by means of reference to a part of their body, as in expressions such as “a safe pair of hands,” or as in

In addition to entities, processes are often transformed in linguistic representation. Actions and processes are described as abstract entities or states, for instance. SAA also provides a systematic distinction of these types that goes beyond considering the verbs of sentences (Fig 9). Fundamentally, processes may be

narrative or conceptual; conceptual processes include classifications ('He is a refugee'), analytical ('This group of people includes A, B and C') and symbolic ('She is the head of the family'). In English, many conceptual processes use 'be', 'have' and other function verbs.

In contrast, narratives may be agentive (specifically actional, reactional, mental/emotional or verbal) or non-agentive. Non-agentive narratives do not include an animate, sentient character but represent events as a natural process without volition (evolution, the tides, a flood, the sunrise). Agentive narrative processes include an acting entity capable of volition; such process may be simple actions (eating, sleeping, travelling), reactional (witnessing, seeing, experiencing), mental or emotional (thinking, feeling, considering, doubting, fearing, worrying) or verbal (saying, writing, shouting). These distinctions are sometimes not clear in actual language use, identifying ambiguity is then part of our analysis.

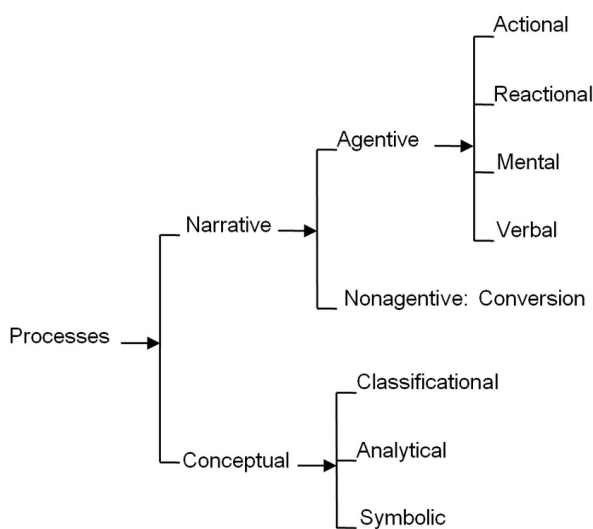


Fig. 9: Process types

Analysing the process types that the characters of key narratives are engaged in can indicate if, for instance, migrants are predominantly presented as having agency or not; as acting or reacting; as feeling/thinking; or as having a voice/articulate. Outside narrative processes, we would find migrants in conceptual process types, for instance in statistics. These may be classifications, analysis or symbolic.

We propose that the qualitative in-depth analysis use the categories of Fig. 8 and 9 to discuss how the main elements of narratives surrounding irregular migration are represented across the three domains, politics, media and civil society – both inside and outside explicit narrative structures.



### 3. References

- Baker, Paul et al. (2008): A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society* 19(3): 273-306.
- Baker, Paul/McEnery, Tony (2005): A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics* 4:2 (2005), 97–226.
- Leeuwen, Theo van (1995). Representing Social Action. *Discourse and Society* 6(1):81–106.10.1177/0957926595006001005
- Leeuwen, Theo van (1996). The Representation of Social Actors. In C. R. Caldas-Coulthard and M. Coulthard (Eds.), *Texts and Practices: Readings in Critical Discourse Analysis* (pp. 32–70). London: Routledge.
- Leeuwen, Theo van (2008). Chapter 2 Representing Social Actors. In *Discourse and Practice: New Tools for Critical Analysis*, Oxford Studies in Sociolinguistics. <https://doi-org.uaccess.univie.ac.at/10.1093/acprof:oso/9780195323306.003.0002>, accessed 31 Jan. 2024.
- McEnery, Tony/Gabrielatos, Costas (2006): English Corpus Linguistics. The Handbook of English Linguistics, edited by Bas Aarts, April McMahon. London: Blackwell.
- Orpin, Debbie (2005): Corpus Linguistics and Critical Discourse Analysis Examining the ideology of sleaze. *International Journal of Corpus Linguistics* 10:1, 37-61.
- Rheindorf, Markus (2019a): Chapter 2: Working with corpora large and small. In *Revisiting the toolbox of Discourse Studies. New trajectories in methodology, open data and visualization* (pp. 21-85). London: Palgrave Macmillan.
- Rheindorf, Markus (2019b): Chapter 7: Case Study 3: Retold narratives of Holocaust Survivors and their Children. In *Revisiting the toolbox of Discourse Studies. New trajectories in methodology, open data and visualization* (pp. 254-279). London: Palgrave Macmillan.
- Rheindorf, Markus (2019c): Chapter 9: Case Study 3: Media Representations of Refugees, Asylum Seekers, Immigrants and Migrants. In *Revisiting the toolbox of Discourse Studies. New trajectories in methodology, open data and visualization* (pp. 307-332). London: Palgrave Macmillan.
- Rheindorf, Markus/Wodak, Ruth (2018): Borders, Fences, and Limits—Protecting Austria from Refugees: Metadiscursive Negotiation of Meaning in the Current Refugee Crisis. *Journal of Immigrant and Refugee Studies* 16/1-2, 15-38.
- Vollmer, Bastian (2021): *Borders Revisited – Discourses on the UK Border*. Basingstoke & New York: Palgrave Macmillan.
- Wodak, Ruth/Rheindorf, Markus (2017): Whose story?—Narratives of persecution, flight and survival told by the children of Austrian Holocaust survivors. In: De Fina, Anna et al. (eds.): *Linguistic Superdiversity*. Washington, D.C.: Georgetown University Press, 17–36.

## i-CLAIM Consortium



Utrecht  
University



UNIVERSITY OF  
BIRMINGHAM



Ca' Foscari  
University  
of Venice  
Department of Philosophy  
and Cultural Heritage



UNIVERSITY  
OF WARSAW



HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI



Katholische  
Hochschule Mainz  
Catholic University  
of Applied Sciences



TOGETHER FOR  
JUSTICE & EQUALITY  
THE JOINT COUNCIL  
for THE WELFARE  
OF IMMIGRANTS

**act:ionaid**  
— REALIZZA IL CAMBIAMENTO —

*Centrala*

**ASSOCIATION  
FOR LEGAL  
INTERVENTION**

CONFEDERATION  
**SYNDICAT  
EUROPÉEN  
TRADE UNION**



Deaconess  
Foundation

KATHOLISCHES  
FORUM  
LEBEN IN DER  
**LEGALITÄT**

**FAIRWORK**

## Contact info

[icclaim@uu.nl](mailto:icclaim@uu.nl)

**For press inquiries:**

I-CLAIM Communications Manager

[miriam.mir@ceps.eu](mailto:miriam.mir@ceps.eu)

## Follow us



[www.i-claim.eu](http://www.i-claim.eu)



Funded by  
the European Union



UK Research  
and Innovation

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or UK Research and Innovation. Neither the European Union nor the granting authority can be held responsible for them.

# I-CLAIM

Improving the Living  
and Labour Conditions  
of Irregularised Migrant  
Households in Europe