



UNIVERSITY
OF WARSAW



FACULTY OF
ECONOMIC SCIENCES

WORKING PAPERS

No. 23/2026 (517)

USING TWEET PHRASES TO PREDICT THE TRANSITION FROM MIGRATION INTENTIONS TO REAL-LIFE MOVEMENTS

ANNA JANICKA
HAYK AMIRKHANYAN

WARSAW 2026

ISSN 2957-0506



Using tweet phrases to predict the transition from migration intentions to real-life movements

Anna Janicka^{1}, Hayk Amirkhanyan¹*

¹ *University of Warsaw, Faculty of Economic Sciences*

* *Corresponding author: ajanicka@wne.uw.edu.pl*

Abstract: Understanding the transition from migration intentions to actual movements remains a challenge in demographic research, often hampered by a persistent intention-behavior gap. This research examines how linguistic variations in individuals' stated migration intentions, particularly on social media, help predict actual movements. We hypothesize that the specific phrasing of the intention reflects the strength and proximity to realization. To verify this claim, we compiled a unique dataset based on migration-related tweets from Polish users posted between 2006 and 2021, identifying potential migrants via Polish keywords that indicated some form of migration intent. The collected statements were manually categorized for linguistic nuances and contextual factors. Actual migration was subsequently tracked by detecting significant country changes (minimum 30 or 180 days) in users' geotagged tweet histories, and logistic regression models were applied to assess the predictive power of different formulations. Our findings demonstrate that the linguistic phrasing of migration intentions correlates significantly with their realization: users employing “planning” keywords were most likely to migrate, while those expressing mere “preferences” were the least likely. Additionally, intentions communicated through multiple tweets, as well as those linked to concrete goals, such as education, positively predicted actual movement. Conversely, tweets with negative contexts, hypothetical conditions, or perceived obstacles were related to a reduced actual migration probability. This study empirically confirms that subtle linguistic variations and contextual elements within spontaneous social media expressions may be treated as indicators of migration intention strength and its likelihood of translating into actual behavior. Our results highlight the importance of differentiating intention forms to enhance migration predictions and refine future survey designs.

Keywords: migration intentions, migration behavior, migration decision-making, intention-behavior gap, social media analysis, Twitter data, geolocated data, linguistic cues, migration prediction, human mobility, Poland

JEL codes: J61, D91, J10, C55, D83

Funding: The research was funded by the National Science Center (Narodowe Centrum Nauki <https://www.ncn.gov.pl/>) in Poland, grant number 2018/31/D/HS4/01523

1. Introduction

Understanding and predicting migration behavior remain central objectives in demographic research and policy analysis. A common empirical strategy involves measuring individuals' stated intentions to migrate, typically using survey instruments, and treating these declarations as predictors of future mobility (Docquier et al., 2014; Migali & Scipioni, 2019; Tjaden et al., 2019). This approach is based on established behavioral theories, particularly the Theory of Planned Behavior (TPB, Ajzen, 1991), which treats intentions as immediate antecedents of behavior.

However, a consistent empirical finding is that migration intentions translate into actual movements only to a limited extent. Longitudinal studies indicate that only a minority of individuals who express an intention or plan to migrate ultimately do so (e.g., Creighton, 2013; van Dalen & Henkens, 2013), leading to an intention–behavior gap. In general, stated intentions may fail to materialize due to changing preferences, external constraints, or limited behavioral control; in the context of migration, where mobility decisions are embedded in complex institutional, economic, and social environments that constrain individual agency (Carling, 2002, 2014; de Haas, 2021), this gap is particularly pronounced.

Even though the intention–behavior gap in migration research is well documented, the mechanisms leading to actual migration seem understudied. In particular, existing research in this field typically treats migration intentions as relatively homogeneous constructs, despite substantial variation in how individuals conceptualize and articulate prospective mobility. Survey-based approaches rely on specific question formulations, which capture different dimensions of future-oriented thinking, including aspirations, preferences, willingness, expectations, and plans (Carling & Mjelva, 2021). These categories are often used interchangeably in empirical analyses, even though they represent qualitatively different levels of commitment and proximity to action.

This lack of differentiation raises a fundamental measurement problem. If expressions such as “thinking about moving,” “wanting to move,” and “planning to move” correspond to distinct stages in the decision-making process, then treating them as equivalent obscures variation in the likelihood of realization. From a theoretical perspective, such variation can be interpreted in terms of differences in intention strength, stability, and perceived behavioral control (Ajzen, 1991; Ajzen & Kruglanski, 2019). Stronger, more specific, and more action-oriented intentions are more likely to translate into behavior than weak or hypothetical considerations. However, even though theoretical grounds exist (Carling & Mjelva, 2021),

empirical evidence linking the formulation of migration intentions to their realization remains scarce.

This paper aims to fill this gap by examining how individuals express migration intentions and how these expressions relate to subsequent migration behavior. Our analysis is based on social media data, which has increasingly been used in research on human mobility patterns and spatial behavior (e.g., Hawelka et al., 2014). However, the potential of social media platforms for studying the transition from intention to action remains underexplored.

In our study, we argue that linguistic variation in migration-related statements reflects differences in intention strength as well as the stage within the decision process. Expressions indicating planning or concrete preparation may be interpreted as signals of stronger, more proximal intentions, whereas expressions of desire or consideration reflect weaker, less committed states. Basing on existing theories of the intention–behavior transition, we argue that this dimension is crucial for understanding which intentions are ultimately realized.

For empirical verification, we construct a dataset of migration-related tweets posted by users in Poland and track their subsequent geographic movements. We analyze spontaneous statements, which allows to observe how intentions are formulated in natural contexts, and not as responses to predefined survey questions. The framework permits us to directly link the content and formulation of intention statements to subsequent migration behavior.

The contribution of this paper is threefold. First, it advances the theoretical discussion of migration decision-making by explicitly linking the realization of intentions to their wording and strength. Second, it introduces a novel approach that captures heterogeneity in intentions beyond what could be observed in survey data and allows to assess how phrasing of the migration desire and its context influence the accuracy of migration behavior predictions. Third, it provides grounds for the interpretation and design of migration intention measures, underlining the importance of different levels of commitment.

2. Theoretical background: migration intentions vs migration behavior

It is common to treat migration as a two-step process, where the first step is the formation of a thought to migrate, and the second is a conversion of this thought into actual migration (Carling & Mjelva, 2021). Various researchers use different names to describe the thoughts or feelings towards migration in the first step, such as: aspirations (Carling, 2002), intentions (Becerra et al., 2010; van Dalen & Henkens, 2008), desires (Collins, 2019), plans (Epstein

& Gang, 2006), expectations (Carling & Pettersen, 2014) etc. In what follows, we will use the term “intention” as an umbrella term for all the concepts related to the first phase of the process unless clear differentiation is necessary.

Different terms are also used to describe the transition process of the second stage. According to the theory of reasoned action (Ajzen & Fishbein, 1980), behavioral intentions are both the immediate determinants and the best predictors of behavior. Ajzen (Carling & Mjelva, 2021)(1991) in the TPB stresses the importance of what he calls the perceived behavioral control in this transition. This is an element related to behaviors which are not completely under volitional control – as in the case of migration. The more control, the more likely the behavior, *ceteris paribus*. In the migration decision context, where the control may be on the lower end, this means that intentions will, in all likelihood, not be perfect predictors of behavior.

In similar frameworks, adjusted to the migration context, Carling (2002, 2014) uses the notion of “ability” and de Haas (2021) the concept of “capability” in place of perceived behavioral control. Within Carling’s framework, the strength of aspirations should be considered as a continuum, from “clearly wishes to stay” to “clearly wishes to move”. Only individuals with attitudes near the extremes are expected to provide predictable answers to questions about intentions; for the individuals from the middle segment, responses may depend on the context. However, even if answers do relate to actual intentions, this does not yet signify automatic fulfillment of declarations, even in cases when migrants are able to do so. This happens for several reasons.

First of all, in order to be able to translate migration intention into action, time must pass. During this period, a lot may happen in an individual’s life. Ajzen (1991) underlines that both intentions and perceived behavioral control must remain stable in the interval between assessment and the occurrence of the behavior in order for the intentions to translate to actions. If migration intentions are changed or abandoned, realization of previous declarations will not occur.

Second, migration might be thought of as a means to an end – fulfilling a more general goal. As it comes to goals in the transition from intentions to actions, Toth-Bos et al. (2019) consider them in particular in the pre-migration stage. This means that for the transition to occur, goals should not vary. Meanwhile, goals are unstable and perennially fluctuating (Ajzen & Kruglanski, 2019).

However, intention or goal changes are not the only reason for lack of fulfillment of migration intentions for capable migrants. A crucial element is the “measurement” precision of the intentions. There are many ways to phrase a survey question about migration in the future. Carling and Mjelva (2021) place what they call “the description of the mindset” at the core of the different approaches. They identify eight general categories of specifications: 1) consideration, 2) preference, 3) willingness, 4) necessity, 5) planning, 6) intention, 7) expectation, and 8) likelihood. In other words, one may ask about

- intentions: “Do you intend to move...”(De Jong et al., 1985; Gardner et al., 1985),
- plans: “Are you planning to move...”(Tjaden et al., 2019; van Dalen & Henkens, 2013),
- willingness to go abroad: “Would you be willing to...”(Efendic, 2016; Epstein & Gang, 2006; Fidrmuc & Huber, 2007),
- thoughts about moving: “Have you ever thought of moving...” (Creighton, 2013; De Jong, 2000),
- preferences for location: “Would you like to live in...” (Becerra et al., 2010),
- pressure: “I must migrate in order to...” (Becerra et al., 2010),
- expectations: “I see myself in...”(Cairns & Smyth, 2011), and so on.

Obviously, for the same individual, the responses to questions about potential movement might vary, depending on the phrasing. A person might wish to move, would like to go abroad, but at the same time does not plan to do so in a specific period of time. Some formulations are tentative and exploratory, such as “thinking about leaving” or “considering going abroad.” Others imply greater commitment and temporal proximity, such as “planning to move” or “I will move next year.” Still others refer to desirability, necessity, or expectation rather than intention in a strict sense. These differences correspond to different stages in the decision process and therefore to different probabilities of realization.

Linguistic variation is indicative of variation in intention strength, stability, and proximity to implementation. Gödri and Feleky (2017) underline that capturing an “insufficiently serious” migration intention may underlie the observed lack of plan realization; desires and dreams are going to be less reliable for predictions than actual migration intentions. In a similar vein, Carling and Mjelva (2021) consider preference, willingness and necessity (an “extreme form of preference”) to be evaluative mindsets, intention and planning to be representing action-oriented statements (which should therefore lead to higher transition rates to actual behavior), while expectation and likelihood to be the most hypothetical. The crucial issue is therefore not

merely whether an individual expresses an intention to migrate, but how that intention is formulated and how close it lies to a concrete plan of action.

Other intention phrasing aspects which, according to Carling and Mjelva (2021), should be considered due to their effect on intention measurement, include the timeframe (whether the declarations refer to the time of the interview or past considerations), spatiality (whether questions are about staying or leaving), duration of the move or the conditionality (conditional or hypothetical nature of questions). Furthermore, studies where a specific context (for example, going abroad to work or to a specific country) for migration is considered are bound to differ from studies where migration is treated more generally.

What needs to be underlined is that even though researchers agree that different formulations measure different aspects of the first phase in the two-step migration process, existing studies do not explore the effects of the measure used on the reliability of behavior predictions for the second stage. Studies of intentions where more than one formulation is applied are scarce, and in most cases do not go beyond modelling the first step. This is for example the case for Steiner (2019), who analyzed the first-stage remigration intentions and plans for immigrants in Germany. The author found that the two constructs are triggered by different factors: intentions by “wishful thinking” or a desire to live elsewhere, while plans are based on concrete events mainly related to the professional and personal life course. The analysis by Wanner (2021) used the same data and went a step further by also considering the actual moves. However, even though the survey included a series of questions related to preferences, considerations and plans, he only used the preference measure for predicting migration behavior (possibly because questions were conditional, e.g. the question about plans was asked only if a respondent had considered migration).

In a different setting, Klöble (2021) analyzed the determinants of three dimensions: desire, plan and preparation for migration in a multi-country setting based on the Gallup Poll. She found evidence that the more concrete the migration intentions become, the more potential migrants are self-selected. However, she did not consider actual migration behavior. On the other hand, in a study of Spanish students, Hoppe and Fujishiro (2015) did not explicitly study wording, but went beyond the two-phase migration model. Those who intended to migrate were divided into predecisional, preactional and actional phases, depending on additional actions taken (such as having discussions with others, obtaining information, making logistical arrangements etc.). The authors found that being in the preactional and actional phases predict (although not perfectly) the actual movement, while being in the predecisional phase does not.

To the best of our knowledge, apart from the studies mentioned above, there are no other empirical approaches which would simultaneously consider different intention phrases and actual moves. Some additional evidence of phrasing relationships may be found by comparing results of different studies, where different formulations are used. In one of the older studies, De Jong (2000) analyzed to what extent “thinking about migration” is a good proxy for migration behavior for the case of internal migration in Thailand. He concluded that the determinants of migration intentions and subsequent migration behavior differed, but nonetheless found some relationships between them. In a study involving Dutch citizens, Van Dalen and Henkens (2013) found that more specific migration intentions were good predictors of actual behavior, with 34% of respondents who had declared a plan to migrate (without a specific time frame) actually emigrating within the follow-up period of 5 years. On the other hand, when aspirations, rather than intentions, of inhabitants of Mexico were studied, the relationship between thoughts and actions was more limited, with only 11% of aspirants to international migration following through with any type of migration during the 3-year period of observation (Creighton, 2013).

Even though there are notable differences in the realization fractions in the above-mentioned studies, the country contexts differ significantly, which weakens the argument of the importance of intention phrasing. We argue that indeed, phrasing is theoretically informative because it indexes the distance between an expressed intention and its possible realization. We propose an approach which allows to juxtapose different wording and actual migration. Instead of using a set of predetermined questions, in our analysis we will look into the transition of migration intentions into actions, considering the way in which the potential migrants themselves formulate their intent. We will do this by studying the intention-behavior transition in the context of social media – posts on the social medium named Twitter at the time of data collection – of Poles expressing the thought about international migration in some form.

For Poland, two-stage migration decision models have not been studied in detail. Basic fractions of intent realization were provided by Czapiński and Panek (2009) for a representative country-wide survey, referring to a question phrased in terms of migration plans (to move abroad to work). The authors mentioned that the majority of migration plans were not translated into actions (86% for the time span 2007-2009). A more detailed analysis, based on the same dataset, was provided by Brzozowski and Coniglio (2021). These authors found that the factors shaping individuals’ intent to migrate diverge substantially from determinants of real migration decisions. The authors attributed the failure to realize migration plans to the differences between

the ex-ante expected costs and benefits of international migration and the actual ones, which are out of the control of those who wish to migrate. They did not, however, explicitly model transition from intention to behavior.

The choice of Poland as a case study allows to talk about migration intentions of inhabitants of a state where (1) emigration is fairly prevalent, allowing for a potentially wide analysis, (2) nationals do not face formal migration constraints for dominating destination countries (EU), allowing for a potentially higher transition rate from intentions to actual migration, and (3) the population is fairly homogenous in terms of language and ethnic backgrounds.

3. Data and Methods

Social media have become a commonly used tool in social studies in general, and in migration studies in particular. Apart from numerous advantages such as data availability, size and costs of obtaining, several drawbacks of social media data have been discussed in the literature. First of all, data are only available for social media users, so it is hard to expect findings to be generalizable to the whole population. For studies of potential migrants, this limitation may not prove to be as important as for other contexts, since a large share of individuals in the most mobile age (25-45 years) have access to and use social media on a daily basis (Firth et al., 2019). Second, relevant data is “contaminated” with non-relevant content, as bots, tourists, and other users who travel frequently may make it harder to study migration flows of “real migrants”. Third, information in social media takes different forms, such as pictures, movies or links; an analysis restricted to text content only (as in our case) is obviously going to be limiting. And, finally, additional information, beyond the text (such as location) is not available for all users. Meanwhile, in the analysis of migration, it is the ability to correctly assess the location of an individual which is of crucial importance.

There are many definitions of migrants based on social media content (see, e.g., Hausmann et al., 2018; W. Huang et al., 2014; Kim et al., 2020; Mazzoli et al., 2020; Zagheni et al., 2014). For instance, Kim et al. (Kim et al., 2020), focus on residence and nationality to identify migrants. In their study, a user’s residence is the place where the user spends most of the time during one year and nationality is predicted using tweet and profile language, as well as social ties. A person is considered a migrant if her place of residence differs from nationality. Huang et al. (2014) also focus on predicting nationalities of twitter users by means of twitter data (hashtags, geography of users’ network, language used, etc.). On the other

hand, Zagheni et al. (2014) define an individual's country of residence (for a specific period of four months) as the country where the Twitter user has the majority of their geotagged tweets. Then, a change in the country of residence in any of the four-month periods under study is considered to be international migration.

In this study we adopt a slightly different approach. We take advantage of the special features of the Polish population: first, immigrants in Poland have, until very recently, been a minority of the population, with migrants from the prevailing source countries (Ukraine and Belarus) frequently not mastering Polish beyond the elementary level (Levchuk, 2021). Second, foreigners who learn Polish without familial motivators tend to be rare (Drozdowski, 2011). Finally, Polish is not spoken in other countries. Hence, for the purpose of this study, we assume that an individual who tweets in Polish is a Pole.

Under this assumption, we first identify those Polish Twitter users who have any tweet about migration intentions posted from anywhere in the world (see detailed procedures below). We then take the subset of users who have any geotagged tweet (1,189 users) and manually check the tweets to identify those containing "real" intentions and to classify them under different categories and contexts. Accordingly, our approach utilizes the tools of big data to obtain a large, rich dataset of potential migrants. However, by manually selecting the real migration intentions we overcome a major shortcoming of big data mentioned earlier: non-relevant content.

For users who are identified as expressing "real" migration intentions, we proceed to check the location of their geotagged tweets (as far back in user's tweet history as available) to identify actual migrants. We define migration based on a "significant change" of the location country, for various time spans. A short-term migrant is a Twitter user for whom we detect a significant change of country, i.e. a change taking place at least 30 days after any previous change, and at least 30 days before the next change (corresponding to seasonal migration). We also consider longer-term migration, with the no-change time span lasting at least 180 days before and after, as well as intermediate time spans as robustness checks.

All data are analyzed after anonymization. Consent from participants was not obtained, as the data were derived from publicly accessible Twitter posts and all identifiable user information (user IDs, tweet IDs, location apart from information about the country), as well as tweet content (after categorization) has been removed. The procedure is approved by the Ethics Committee of the Faculty of Economic Sciences at the University of Warsaw in decision 12/2026.

4. Procedures

In the first step, we define a list of 41 Polish language keywords that potentially indicate migration intentions (e.g., “I plan to leave Poland”, “I intend to emigrate”, “I want to move from this country”, “I think about migrating from Poland”, “I’m moving abroad” etc.; see Supplementary materials for the list of keywords and the number of tweets found). In these keywords, we allow for different phrasing of the migration intention (plan/intention/wish/thought), adjusting the different mindsets identified by Carling and Mjelva (2021) to the Polish language (see Table A1 in the Appendix for a list of these keywords). We do not operationalize the likelihood and expectation categories at this point, but we introduce an “announcement” (“I am leaving”) category that better reflects the performative and declarative nature of Twitter discourse. We also allow for differentiated references to Poland (different attitudes and spatiality). Using Twitter’s API for Academic Research, we perform a full-archive search (run on June 2, 2022, covering the period from March 22, 2006 until December 31, 2021) of each keyword for an exact match, ignoring any re-tweets.

Since generally only 2-3% of tweets on Twitter have location information (Huang & Carley, 2019), at this first step we do not apply any geolocation or country restrictions and assume that tweets come from Poland (indeed, 15 out of 50 keywords include “from Poland” and the rest of the keywords contextually indicate leaving from Poland: e.g., I want to go abroad [from Poland]). In this way we identify 10,518 tweets, coming from 7,956 users (some users have more than one tweet about migration intentions). Depending on the user’s location-sharing preferences, some of the tweets (2.3%) also have a geotag.

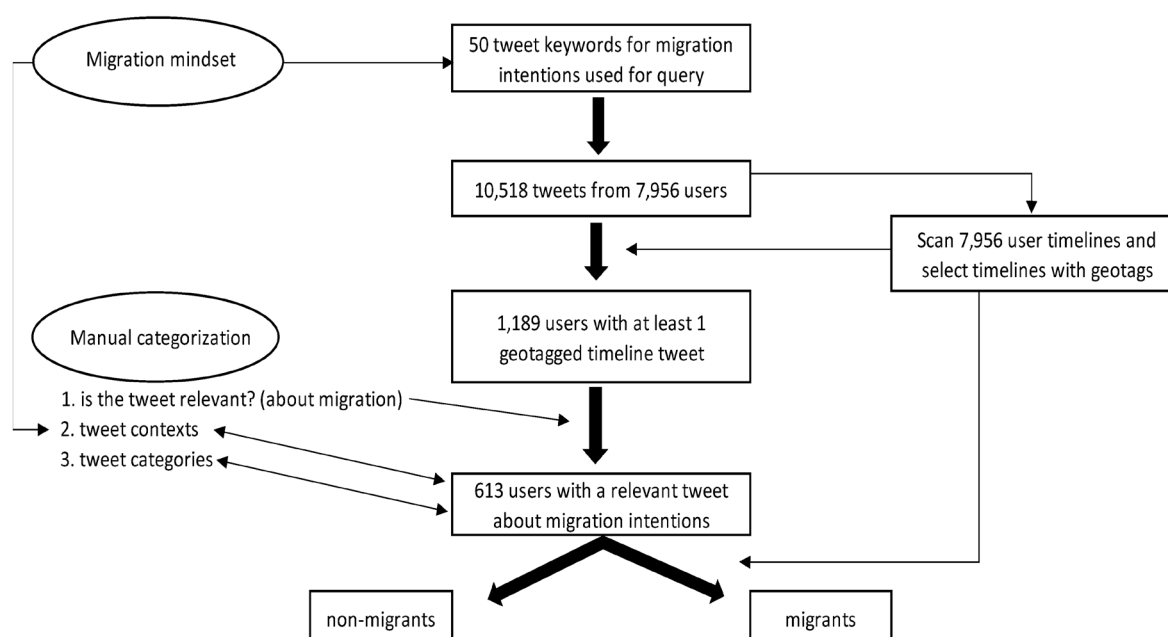
In the second step, we utilize Twitter API’s “user tweet timeline” endpoint that allows to scan the most recent 3,200 tweets of a user. Because location information is of crucial importance for identifying actual migrants, we scan the timelines (i.e., users’ tweets history) of all the 7,956 users recognized in the first step and request only the tweets that have either country or coordinates information. This process is run in December 2022 – January 2023. As a result, we identify a group of 1,189 among the potential migrants who have at least one tweet with a country or coordinates information in their latest 3,200 tweets.

Third, in order to identify the “real” would-be migrants, we manually categorize the tweets with migration intention keywords for these 1,189 users. The manual check is conducted by two researchers independently. In the case of inconsistencies, which were around 5% of the tweets, a consensus is reached by discussing each tweet individually. We drop a tweet if it is not about own migration plans (e.g., is about someone else or suggests short-term travels).

Additionally, for the tweets that are classified as relevant, we manually add several categories and context tags, e.g., planning to leave after high-school or studies, tweet with a negative attitude or context (e.g., complaining about politics), intending to perform work or studies abroad. At the time of categorization, the attempts to use automated models were made but abandoned, since models failed to grasp the context and overtone or determining whether a tweet was related to own migration or migration of others (due to the specificity of the Polish language). The list of categories and context tags is available in Table A2 in the Appendix. For the users that have multiple migration intention tweets, we pick one “representative” tweet (earliest, if all tweets are similar, or the one providing the most details) and additionally record the number of tweets about migration intentions.

Finally, for those users, whose tweets were categorized as indicating intentions to go abroad, we add information from the collected timelines. As a result, our dataset consists of initial migration intention tweets and information derived from geotagged timeline tweets for 613 users. Figure 1 summarizes the process of obtaining the final dataset. Even though during initial data collection we had access to publicly available information that might potentially identify Twitter users, all such information was dropped before analysis, and the final dataset is fully anonymized.

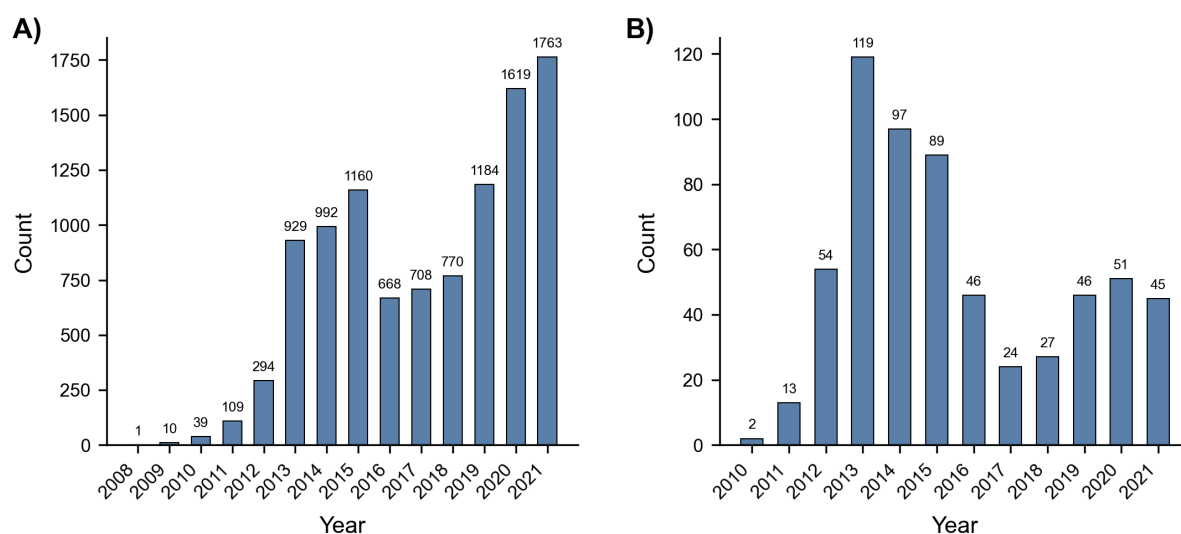
Fig. 1. The process of obtaining the final dataset



5. Data

Figure 2 shows the yearly distribution of the categorized real migration intention tweets of our final 613 users for whom we have geotagged timeline tweets (right panel B)). This distribution is complemented with the overall number of tweets found in the initial keyword search (left panel A)) to show the differences in patterns.

Fig. 2. Number of migration intention tweets per year: A) raw data and B) geotagged users' tweets



Note: “raw data” represents the distribution of overall number of tweets found (tweets dataset), “geotagged” shows the distribution of “real” migration intentions (i.e., after manual categorization) of our final 613 users.

In general, an increasing trend in expressing migration intentions is visible, with the exception of the years 2016-2018, when a significant drop in the number of tweets about leaving Poland was observed. However, the grey bars relate to tweets which were not filtered, and thus contain news tweets such as: “30% of Poles want to go abroad for work #hashtag”. For the hand-picked and categorized tweets, the proportion of relevant intention tweets amounted to roughly 52% (613 relevant out of around 1,189 tweets); the share of relevant tweets in the un-categorized part (i.e., in 9,329 tweets) may be smaller (if we assume that users who migrate are more inclined to share location than the general population of Twitter users). Besides, around 2,000 users have multiple tweets.

As it comes to the tweets from the 613 users with geo-tagged timelines, it seems that 2013-2015 are the years with most tweets about migration plans. It is hard to provide a definite explanation, but it seems plausible that the overall cyclical pattern is related both to changes in the political and economic landscape in the country and to varying location-disclosure policies.

Our final dataset contains information for 613 unique users with matched migration intention and geotagged timeline tweets (142,215 tweets in total). Among those timeline tweets, 16% had exact coordinates. For the tweets that had coordinates but not country information, we utilized the `geopy.geocoders` Python package to convert the coordinates to a country name. For some coordinates no country was found in our coordinates-to-country conversion, and these tweets were omitted. Simultaneously, 99.7% of tweets had country information assigned by default, through “Twitter places”. Table 1 presents the top countries and languages detected in the sample of geotagged timeline tweets. The majority of our sample users twitted predominantly from Poland, in Polish. The top destination countries seem to be UK and US, with English being the second dominant language of the timeline tweets. This distribution seems to be mostly in line with official statistics of Polish emigration, with the exception of a higher share of United States locations. We attribute this to country-specific Twitter usage patterns.

Table 1. Top 10 countries and languages of timeline tweets (all geotagged tweets)

Country	Freq.	Percent	Cum.	language	Freq.	Percent	Cum.
Poland	110,246	77.52	77.52	Polish	107,545	75.62	75.62
United Kingdom	10,041	7.06	84.58	English	19,917	14	89.63
United States	7,908	5.56	90.14	und	5,023	3.53	93.16
The Netherlands	3,526	2.48	92.62	qme	1,415	0.99	94.15
Germany	2,270	1.6	94.22	German	896	0.63	94.78
Italy	790	0.56	94.77	zxx	821	0.58	95.36
Ireland	634	0.45	95.22	Spanish	765	0.54	95.9
Austria	605	0.43	95.65	Tagalog	691	0.49	96.38
Spain	561	0.39	96.04	Indonesian	517	0.36	96.75
Belgium	536	0.38	96.42	Portuguese	468	0.33	97.08

Note that the language of the tweet is detected by Twitter: “und” stands for “undefined”; “qme” means “tweets with media links”; “zxx” means “tweets with either media or Twitter Card only, without any additional text”

It is worth highlighting that our geotagged timeline tweets (coming from the latest 3,200 tweets) are not necessarily “very recent” and cover the years 2011-2023, with the average being 2017. In fact, for 130 users (or 21.2% of our 613 users) the geotagged timeline “starts” earlier than the migration intention tweet. One hundred fifteen users (or 18.8% of 613 users) posted geotagged tweets from abroad before creating a tweet about migration intentions (see Table 2).

Table 2. Summary statistics of key variables (for users)

Variable	Mean	Std. dev.	Min	Max
Number of country changes	2.78	10.26	0	118
Number of geotagged tweets per country	41.05	80.37	1	678
First geotagged tweet in timeline detected abroad (not Poland)	0.188	0.39	0	1
Number of days since intention tweet to first geotagged tweet	566.80	918.55	-1944	4223

It is also visible from data presented in Table 2 that the geotagged timelines on average “start” 567 days (1.55 years) after the migration intention tweet. Our users on average changed the country location close to three times within the recorded timeline and created geotagged tweets around 40 times per country they stayed in. A country change is calculated based on all the timeline tweets (the first tweet of the day) of a particular user, by detecting the country changes in relation to the tweet location of the previous date. Surely, this may involve short travels, hence we focus on “significant” country changes: mobility for more than 30 or 180 days. In the analysis that follows, we do not distinguish between one-time and repeat migrants, but only concentrate on whether a person migrated at any point in time after the migration intention tweet.

In the final database, there are 183 (29.85%) users with a minimum 30-day country change and 123 (20.1%) with a minimum of 180 days change (see Table 3). One hundred seven users posted multiple migration intention tweets. Table 3 also summarizes the occurrence of categories and contexts identified in the process of manual categorization, as well as the statistics of keyword categories. The most common category of tweets consisted of general statements (80.4%), but migration intention tweets were also likely to include a hypothetical component (23.2%) and/or a negative attitude or context (26.1%). Due to their short form, tweets in general did not include a lot of details. However, 10.8% included an indication of a potential time frame for migration, while in 7.5% of cases a destination country was specified. The most commonly used phrases included “I’m going [abroad/away]” (32.5%), “I will go [abroad/away]” (26.8%) and “I want [to go abroad]” (19.9%). The first two theoretically relate to future actions, but the Polish overtone of these statements is usually hypothetical.

Table 3. Descriptive statistics of users and migration intention tweets

Variable	Frequency	Count
User involved in min 30 days migration	0.299	183
User involved in min 180 days migration	0.201	123
User with multiple intention tweets	0.175	107
Category 1 (giving away stuff)	0.005	3
Category 2 (planning the trip before departure)	0.011	7
Category 3 (leave after studies or high school)	0.160	98
Category 4 (negative attitude or context)	0.261	160
Category 5 (general migration-related statement)	0.804	493
Category 6 (indication of a goal)	0.039	24
Category 7 (obstacles/hypothetical)	0.232	142
Context 1 (work-related)	0.059	36
Context 2 (study-related)	0.020	12
Context 3 (other details)	0.024	15
Context 4 (friend, family or relative in destination country)	0.008	5
Context 5 (destination location specified)	0.075	46
Context 6 (time frame indication)	0.108	66
Keyword category Consideration	0.072	44
Keyword category Preference	0.201	123
Keyword category Necessity	0.056	34
Keyword category Planning	0.025	15
Keyword category Intention	0.271	166
Keyword category Announcement	0.377	231
Keyword category “from this country”	0.210	129

Note: As an example, a tweet “When I finish school, I plan to go abroad to the USA and get my bachelor’s degree there. My uncle lives in Chicago, so he could help me” would be categorized under categories 3 (after school) and 6 (goal) and contexts 2 (study), 4 (relative) and 5 (destination specified). Such a tweet would be under the Planning keyword category. Keyword categories (apart from “from this country”) are mutually exclusive; content and context categories may overlap.

Tweets from the prevailing keyword categories were predominantly categorized as general statements, i.e. without details provided (67%-87%), and statements with a negative context, for example expressing dissatisfaction with the political scene (7%-52%, depending on additional context words). Moreover, a large share of these tweets additionally had a conditional or hypothetical component (up to 53%), e.g. if they included statements such as “I will leave this country if XXX wins the elections”.

6. Results

Figures 3A -3B present the migration rates for minimum 30-day migration and minimum 180-day migration for keyword categories from migration-intention tweets, as well as for manually coded categories and contexts. We find that those who used words related to planning were the most likely to proceed to migrate. On the other hand, those who only expressed preferences for migration (“I want to migrate”) were the least likely to do so. Also, those who used the term “from this country”, rather than “Poland”, were less likely to leave. Other instances of negative context also did not lead to increased short-term mobility. Overall, if migration were to be “running away from something”, it seems to be less likely to be put in place. This result is in line with Ajzen and Kruglanski (2019), who underline the importance of goals for transitioning from an attitude to an intent and action; being disappointed with surroundings and expressing a wish to leave them is not necessarily connected with any active goals an individual might have.

Fig. 3A. Frequency of migration for over 30 days, for users with selected types of tweets

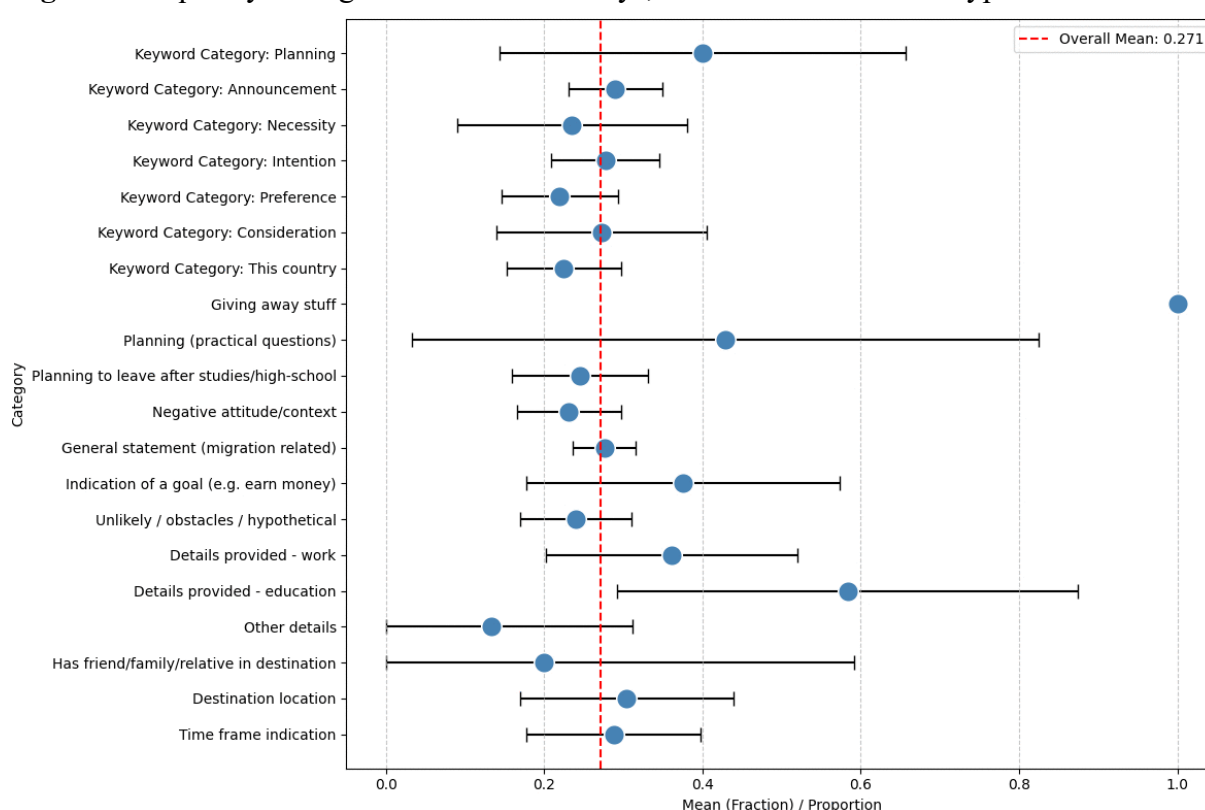
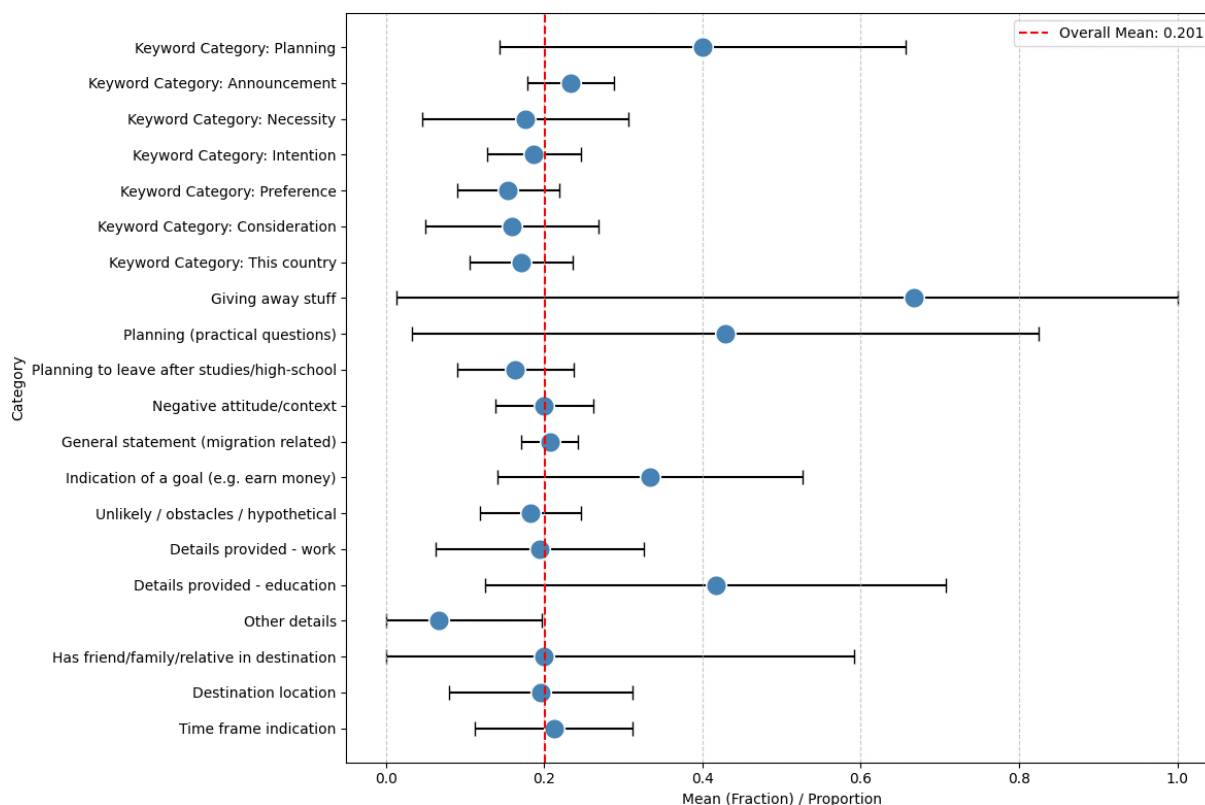


Fig. 3B. Frequency of migration for over 180 days, for users with selected types of tweets

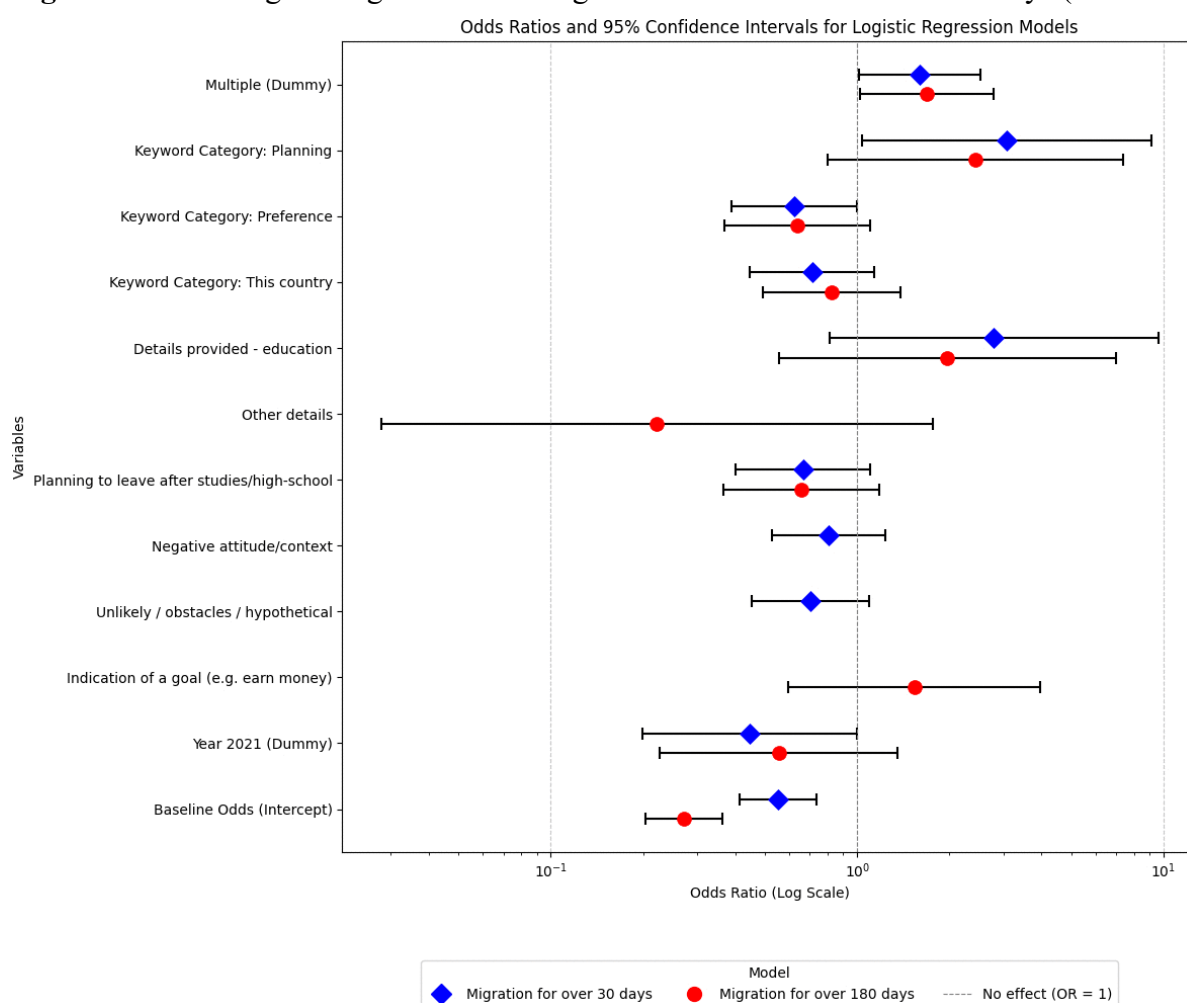
The same may be said for those who expressed some kind of necessity (“I must”), although in this case, the tweet count is small and these tweets frequently not signal any economic necessity, but rather dissatisfaction with the political scene (such as “I must leave this country, I hate the atmosphere and the people around me”). On the other hand, including specific details related to education or work, or other goals, pushed the likelihood of migrating upwards. Unfortunately, the brevity of tweets did not allow to capture the full context and assign potential goals to all tweets (or determine there are none involved).

Our results therefore support conclusions which could be drawn from our brief review in section 2, where researchers using “plans” (such as van Dalen & Henkens, 2013) observed higher transition probabilities than those who studied preferences (such as Creighton, 2013). They also seem to be in line with the suggestions of Carling and Mjelva (2021), who refer to the “plan” phrasing as being already “a step in the direction of behavior”. Furthermore, even though the effects are not statistically significant, the results also align with the observations of Hoppe and Fujishiro (2015), as the statement which may be considered actional (such as giving away stuff) or preactional (practical matters, details of educational-migration plans) coincide with higher probability of migration fulfillment.

Additionally, individuals who had persistent plans and tweeted about migration intentions more than once were more likely to leave (frequency of 38.3% for stays over 30 days and 27.1% for stays over 180 days). All data, not only for 30- and 180- day leaves but also intermediate 60- and 90-day timeframes, are provided in Table A3 in the Appendix.

In order to verify the combined effect of the above characteristics, we estimated logistic models for two alternative dependent variables: taking on a value of 1 if a user migrated for at least 30 days and taking a value of 1 if a user migrated for at least 180 days (at any time after the initial migration tweet). The analysis was performed in STATA 16. In the regression models we present here, we only included variables for which the largest effects for the probability of emigration were identified, disregarding categories with very few counts (full model results available in Table A4 in the Appendix). Apart from the contexts described above, we also used a dummy variable for the initial tweet being posted in year 2021 (since for this year, the potential time span for realization of migration plans was significantly shorter than for other users). The results of the logistic regressions are presented in Figure 4.

Fig. 4. Results of logistic regressions for migration for over 30 and over 180 days (odds ratios)



The signs of parameters obtained in the logistic regression are consistent with the descriptive analysis and the theoretical background. In general, the effects are more visible for shorter-term migration. Having multiple tweets about migration intentions, which is an indication that the preferences are time-consistent, increases the probability that a user will engage in migration. Utilizing the “plan” keyword, which Carling and Mjelva (2021) identified as being already a step towards realization of intentions, also increases this chance, and has the largest effect. On the other hand, using keywords related to preferences (“want”) lead to lower likelihood of mobility. An educational context of the potential leave is more likely to have higher chances of realization, although tweets where (apparently younger) users indicated that the plans relate to a more distant future (after finishing school/studies) led to lower chances of realization within the specified timeframe.

Using negative contexts (the “this country” keyword or others) decrease the probability of migration. The same may be said for specifying some potential obstacles or doubts (or, in the wording of the TPB, indicating a low perceived behavioral control), for short-term migration. Finally, the technical aspect of posting in 2021 (resulting in potentially censored observations) proved to lower the chances of mobility.

Although our results are in line with expectations, what should be underlined is that the overall quality of the logistic models and their predictive power is quite low. This is partially due to the specifics of the variables used in the model – all of them are categorical, none are continuous. Low predictive properties are not uncommon when studying migration decisions even in more favorable settings. The model of De Jong et al.(1985) for predicting actual movements has an R^2 of 0.17, while that of van Dalen and Henkens (2013) – 0.19 (other cited authors did not publish information about model fits). Their results were obtained with a large set of socio-demographic and economic characteristics considered in the models. Meanwhile, in our models no individual characteristics or variables other than those related to the tweet content were included in the analysis. Therefore, we do not control for any user characteristics which are known to influence the process of transition from intent to action.

7. Discussion

Our analysis of the “predictability” of mobility based on intention phrasing revealed in tweets shows that indeed, the linguistic layer translates to observable differences, and certain words and contexts impact the likelihood of putting intentions into actions. In particular, individuals who refer to plans are more likely to move than those who just express a wish.

Although not statistically significant in all cases, the observed differences between keyword categories are in line with the claim that phrasing depicts the distance between the intention and its realization, with some statements being truly “more hypothetical” than others. Additionally, our results suggest that identifying negative or conditional contexts of the inclination may indicate lower chances of actual mobility, while linking intentions to specific goals is connected to higher realization. Overall, based on natural language, we provide a new dimension to the concept of intention strength within the context of migration. We deliver an empirical verification of what has, until now, been justified more by common sense than by rigorous research.

We must admit that this study has certain limitations. Manual categorization of content may be better than machine algorithms, but it’s not totally error-resistant. Due to the brevity of the tweets, some of them may have been mis-categorized. For example, those who tweeted about leaving Poland may have meant not migration, but short-term travel. Such cases should have been filtered out, but it can’t be ruled out that some have been missed.

Additional sources of potential error are related to localization. First of all, some true migrations may have been overlooked in the data. This would happen if individuals did not use geo-tagging for migration, but used geo-tagging for travel, or used geo-tagging at some other point in time when they stayed in Poland. In such a case, they would be treated in our database as non-migrants, even if in fact they did migrate. Also, true movers may have been unnoticed when applying time constraints, especially for longer migration. This might occur if a person did in general change location, but within a 180-day period was also involved in a lot of short-term travel, in which case the main change of location could go unnoticed.

Finally, a source of potential error is that not all users use geo-tagging. Of the initial number of 10,518 tweets with migration intention keywords, less than twelve hundred could be assessed due to the fact that the authors of the remaining tweets never used geo-location. The patterns of the popularity of geo-location are likely to change with time or with changes in data-protection policies. Furthermore, one cannot neglect the fact that there may be positive selection to geo-tagging for users who migrate. Due to our study design we are not able to control for selectivity with traditional variables, since we do not have any information about user characteristics.

Obviously, an analysis of Twitter content related to migration intentions is not representative of the whole population, and not even of the social media users (some of them might have made a migration-related announcement not on Twitter, but for example on

Facebook). However, to the best of our knowledge this is the first attempt to use natural language of potential migrants to study the relationship between migration intention phrasing and the ensuing moves (or lack thereof). This naturalistic, unobtrusive observation afforded by social media is a key strength that bypasses the reactivity or social desirability bias often present in self-reported survey data. Furthermore, we feel that even if not representative, the analysis does push forward our understanding of the importance of intention phrasing for assessing the process of transition from intentions to migration behavior. It also sheds more light on how we should approach studying survey-based declarations for forecasting actions, in particular what questions to ask or which keywords to look for in order to obtain more predictive assessments.

References

- Ajzen, I., 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes, Theories of Cognitive Self-Regulation*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Ajzen, I., Fishbein, M., 1980. *Understanding attitudes and predicting social behavior* (Pbk. ed). Prentice-Hall.
- Ajzen, I., Kruglanski, A.W., 2019. Reasoned action in the service of goal pursuit. *Psychological Review*, 126(5), 774–786. (2019-41283-001). <https://doi.org/10.1037/rev0000155>
- Becerra, D., Gurrola, M., Ayón, C., Androff, D., Krysik, J., Gerdes, K., Moya-Salas, L., Segal, E., 2010. Poverty and Other Factors Affecting Migration Intentions among Adolescents in Mexico. *Journal of Poverty*, 14(1), 1–16. <https://doi.org/10.1080/10875540903272801>
- Brzozowski, J., Coniglio, N., 2021. International migration and the (un) happiness push: Evidence from Polish longitudinal data. *International Migration Review*, 55(4), 1089–1120.
- Cairns, D., Smyth, J., 2011. I wouldn't mind moving actually: Exploring Student Mobility in Northern Ireland. *International Migration*, 49(2), 135–161. <https://doi.org/10.1111/j.1468-2435.2009.00533.x>
- Carling, J., 2002. Migration in the age of involuntary immobility: Theoretical reflections and Cape Verdean experiences. *Journal of Ethnic and Migration Studies*, 28(1), 5–42. <https://doi.org/10.1080/13691830120103912>
- Carling, J., 2014. The role of aspirations in migration. Paper presented at Determinants of International Migration, International Migration Institute, 23–25 September, Oxford.
- Carling, J., Mjelva, M.B., 2021. Survey instruments and survey data on migration aspirations (2; QuantMig Project Deliverable). University of Southampton. <https://www.prio.org/publications/12661>
- Carling, J., Pettersen, S.V., 2014. Return Migration Intentions in the Integration–Transnationalism Matrix. *International Migration*, 52(6), 13–30. <https://doi.org/10.1111/imig.12161>
- Collins, F.L., 2019. Desire as a theory for migration studies: Temporality, assemblage and becoming in the narratives of migrants. In *Aspiration, Desire and the Drivers of Migration*. Routledge.

- Creighton, M.J., 2013. The role of aspirations in domestic and international migration. *The Social Science Journal*, 50(1), 79–88. <https://doi.org/10.1016/j.soscij.2012.07.006>
- Czapiński, J., Panek, T., 2009. Diagnoza społeczna 2009. Warunki i Jakość Życia Polaków, 2.
- de Haas, H., 2021. A theory of migration: The aspirations-capabilities framework. *Comparative Migration Studies*, 9(1), 8. <https://doi.org/10.1186/s40878-020-00210-4>
- De Jong, G.F., 2000. Expectations, gender, and norms in migration decision-making. *Population Studies*, 54(3), 307–319. <https://doi.org/10.1080/713779089>
- De Jong, G.F., Root, B.D., Gardner, R.W., Fawcett, J.T., Abad, R.G., 1985. Migration Intentions and Behavior: Decision Making in a Rural Philippine Province. *Population and Environment*, 8(1/2), 41–62.
- Docquier, F., Peri, G., Ruysen, I., 2014. The Cross-country Determinants of Potential and Actual Migration. *International Migration Review*, 48(s1), S37–S99. <https://doi.org/10.1111/imre.12137>
- Drozdowski, D., 2011. Language Tourism in Poland. *Tourism Geographies*, 13(2), 165–186. <https://doi.org/10.1080/14616688.2011.569569>
- Efendic, A., 2016. Emigration intentions in a post-conflict environment: Evidence from Bosnia and Herzegovina. *Post-Communist Economies*, 28(3), 335–352. <https://doi.org/10.1080/14631377.2016.1166800>
- Epstein, G.S., Gang, I.N., 2006. The Influence of Others on Migration Plans. *Review of Development Economics*, 10(4), 652–665. <https://doi.org/10.1111/j.1467-9361.2006.00340.x>
- Fidrmuc, J., Huber, P., 2007. The willingness to migrate in the CEECs evidence from the Czech Republic. *Empirica*, 34, 351–369. <https://doi.org/10.1007/s10663-006-9028-6>
- Firth, J., Torous, J., Stubbs, B., Firth, J.A., Steiner, G.Z., Smith, L., Alvarez-Jimenez, M., Gleeson, J., Vancampfort, D., Armitage, C.J., Sarris, J., 2019. The “online brain”: How the Internet may be changing our cognition. *World Psychiatry*, 18(2), 119–129. <https://doi.org/10.1002/wps.20617>
- Gardner, R.W., De Jong, G.F., Arnold, F., Cariño, B.V., 1985. The best-laid schemes: An analysis of discrepancies between migration intentions and behavior. *Population and Environment*, 8(1), 63–77. <https://doi.org/10.1007/BF01263017>
- Gödri, I., Feleky, G.A., 2017. Selection of migrants and realization of migration intentions—Lessons from a panel study (Working Paper 29). *Working Papers on Population, Family and Welfare*. <https://doi.org/10.21543/WP.2017.29>

- Hausmann, R., Hinz, J., Yildirim, M.A., 2018. Measuring Venezuelan emigration with Twitter (Working Paper 2106). Kiel Working Paper. <https://www.econstor.eu/handle/10419/179127>
- Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3), 260–271. <https://doi.org/10.1080/15230406.2014.890072>
- Hoppe, A., Fujishiro, K., 2015. Anticipated job benefits, career aspiration, and generalized self-efficacy as predictors for migration decision-making. *International Journal of Intercultural Relations*, 47, 13–27. <https://doi.org/10.1016/j.ijintrel.2015.03.025>
- Huang, B., Carley, K.M., 2019. A large-scale empirical study of geotagging behavior on Twitter. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 365–373. <https://doi.org/10.1145/3341161.3342870>
- Huang, W., Weber, I., Vieweg, S., 2014. Inferring nationalities of Twitter users and studying inter-national linking. *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, 237–242. <https://doi.org/10.1145/2631775.2631825>
- Kim, J., Sîrbu, A., Giannotti, F., Gabrielli, L., 2020. Digital Footprints of International Migration on Twitter, in: Berthold, M.R., Feelders, A., Krempel G. (Eds.), *Advances in Intelligent Data Analysis XVIII* (Vol. 12080, pp. 274–286). Springer International Publishing. https://doi.org/10.1007/978-3-030-44584-3_22
- Klöble, K., 2021. A behavioural perspective on the drivers of migration: Studying economic and social preferences using the Gallup World Poll (Working Paper 4/2021; IDOS Discussion Papers). German Institute of Development and Sustainability (IDOS). <https://doi.org/10.23661/dp4.2021>
- Levchuk, P., 2021. Ukrainian Language in Polish Public Space. *Cognitive Studies | Études Cognitives*, (21), Article 21. <https://doi.org/10.11649/cs.2476>
- Mazzoli, M., Diechtiareff, B., Tugores, A., Wives, W., Adler, N., Colet, P., Ramasco, J.J., 2020. Migrant mobility flows characterized with digital data. *PLOS ONE*, 15(3), e0230264. <https://doi.org/10.1371/journal.pone.0230264>
- Migali, S., Scipioni, M., 2019. Who's About to Leave? A Global Survey of Aspirations and Intentions to Migrate. *International Migration*, 57(5), 181–200. <https://doi.org/10.1111/imig.12617>

- Steiner, I., 2019. Settlement or Mobility? Immigrants' Re-migration Decision-Making Process in a High-Income Country Setting. *Journal of International Migration and Integration*, 20(1), 223–245. <https://doi.org/10.1007/s12134-018-0602-0>
- Tjaden, J., Auer, D., Laczko, F., 2019. Linking Migration Intentions with Flows: Evidence and Potential Use. *International Migration*, 57(1), 36–57. <https://doi.org/10.1111/imig.12502>
- Toth-Bos, A., Wisse, B., Farago, K., 2019. Goal pursuit during the three stages of the migration process. *International Journal of Intercultural Relations*, 73, 25–42. <https://doi.org/10.1016/j.ijintrel.2019.07.008>
- van Dalen, H.P., Henkens, K., 2008. Emigration Intentions: Mere Words or True Plans? Explaining International Migration Intentions and Behavior (SSRN Scholarly Paper 1153985). <https://doi.org/10.2139/ssrn.1153985>
- van Dalen, H.P., Henkens, K., 2013. Explaining emigration intentions and behaviour in the Netherlands, 2005–10. *Population Studies*, 67(2), 225–241. <https://doi.org/10.1080/00324728.2012.725135>
- Wanner, P., 2021. Can Migrants' Emigration Intentions Predict Their Actual Behaviors? Evidence from a Swiss Survey. *Journal of International Migration and Integration*, 22(3), 1151–1179. <https://doi.org/10.1007/s12134-020-00798-7>
- Zagheni, E., Garimella, V.R.K., Weber, I., State, B., 2014. Inferring international and internal migration patterns from Twitter data. *Proceedings of the 23rd International Conference on World Wide Web*, 439–444. <https://doi.org/10.1145/2567948.2576930>

Appendixes

Table A1. List of Polish keywords and their English translations used in the search

Nr	Keyword (PL)	Keyword (EN)	Category	Tweets found	Tweets used
1	Planuję wyjechać za granicę	I plan to go abroad	Planning	59	8
2	Chcę wyjechać za granicę	I want to go abroad	Preference	674	33
3	Chcę wyjechać z Polski	I want to leave Poland	Preference	883	53
4	Chcę wyjechać z tego kraju	I want to leave this country	Preference	344	22
5	Planuję wyjechać z Polski	I plan to leave Poland	Planning	31	1
6	Planuję wyjechać z tego kraju	I plan to leave this country	Planning	3	1
7	Wyjeżdżam za granicę	I am going abroad	Announcement	1203	55
8	Wyjeżdżam z Polski	I am leaving Poland	Announcement	819	77
9	Wyjeżdżam z tego kraju	I am leaving this country	Announcement	759	69
10	Nic mnie tu nie trzyma	Nothing is keeping me here	Announcement	1176	22
11	Czas na emigrację	Time to emigrate	Announcement	499	20
12	Chcę emigrować	I want to emigrate	Preference	583	6
13	Planuję emigrować	I plan to emigrate	Planning	20	0
14	Chcę wyemigrować	I want to emigrate	Preference	288	9
15	Wyjadę za granicę	I will go abroad	Intention	926	72
16	Wyjadę z Polski	I will leave Poland	Intention	846	61
17	Wyjadę z tego kraju	I will leave this country	Intention	425	31
18	Zamierzam wyjechać za granicę	I intend to go abroad	Intention	16	1
19	Zamierzam wyjechać z Polski	I intend to leave Poland	Intention	12	0
20	Zamierzam wyjechać z tego kraju	I intend to leave this country	Intention	7	1
21	Muszę wyjechać za granicę	I must go abroad	Necessity	41	1
22	Muszę wyjechać z Polski	I must leave Poland	Necessity	33	4
23	Muszę wyjechać z tego kraju	I must leave this country	Necessity	32	3
24	Szukam pracy za granicą	I am looking for a job abroad	Announcement	88	10
25	Co zrobić przed wyjazdem za granicę	What to do before going abroad	Announcement	2	0
26	Trzeba wyjechać za granicę	One needs to go abroad	Necessity	48	2
27	Trzeba wyjechać z Polski	One needs to leave Poland	Necessity	66	0
28	Trzeba wyjechać z tego kraju	One needs to leave this country	Necessity	45	2
29	Planuję emigrację	I am planning emigration	Planning	123	3

Nr	Keyword (PL)	Keyword (EN)	Category	Tweets found	Tweets used
30	Planuję wyjazd za granicę	I am planning a trip abroad	Planning	43	2
31	Planuję wyjazd z Polski	I am planning to leave Poland	Planning	18	0
32	Myślę o emigracji z tego kraju	I am thinking about emigrating from this country	Consideration	1	0
33	Myślę o emigracji za granicę	I am thinking about emigrating abroad	Consideration	1	0
34	Myślę o wyjeździe za granicę	I am thinking about going abroad	Consideration	35	3
35	Myślę o wyjeździe z Polski	I am thinking about leaving Poland	Consideration	15	1
36	Myślę o wyjeździe z tego kraju	I am thinking about leaving this country	Consideration	8	1
37	Myślę o emigracji	I am thinking about emigrating	Consideration	268	27
38	Zastanawiam się nad wyjazdem za granicę	I am considering going abroad	Consideration	16	2
39	Zastanawiam się nad emigracją	I am considering emigration	Consideration	47	8
40	Zastanawiam się nad wyjazdem z Polski	I am considering leaving Poland	Consideration	9	1
41	Zastanawiam się nad wyjazdem z tego kraju	I am considering leaving this country	Consideration	6	1
	TOTAL			10518	613

Table A2. List and descriptions of categories and context tags applied in the process of manual categorization.

Category	Category name
1	Giving away stuff
2	Planning (do I need something, practical questions for emigration, looking for a job, etc.)
3	Planning to leave after studies or high-school
4	Negative attitude or context (political views, disliking something, e.g. weather, family, school, town)
5	General statement (migration related)
6	Indication of a goal (e.g. need to earn money)
7	Unlikely / obstacles / hypothetical (e.g. “if something happens I leave”, considering 2 options: leave or something else, language/family difficulties)
Context tags	(Specific details related to the plan / goal)
1	Some details provided - work
2	Some details provided - education
3	Other details
4	Has a friend, family, relative in the destination country
5	Destination location
6	Time frame indication (soon, after a month, next year etc.)

Table A3. Shares of users identified as migrants using 30, 60, 90, and 180 day thresholds, by language keyword categories, manually applied categories and contexts and multiple tweets.

Category	Migration over 30 days			Migration over 60 days			Migration over 90 days			Migration over 180 days			N
	Mean	Lower 95%	Upper 95%	Mean	Lower 95%	Upper 95%	Mean	Lower 95%	Upper 95%	Mean	Lower 95%	Upper 95%	
Overall	0.299	0.262	0.335	0.271	0.236	0.306	0.253	0.218	0.287	0.201	0.169	0.232	613
Keyword Category: Planning	0.6	0.343	0.857	0.4	0.143	0.657	0.4	0.143	0.657	0.4	0.143	0.657	15
Keyword Category: Announcement	0.32	0.26	0.381	0.29	0.231	0.349	0.268	0.211	0.326	0.234	0.179	0.288	231
Keyword Category: Necessity	0.235	0.091	0.38	0.235	0.091	0.38	0.235	0.091	0.38	0.176	0.046	0.307	34
Keyword Category: Intention	0.295	0.226	0.365	0.277	0.209	0.345	0.253	0.187	0.319	0.187	0.127	0.246	166
Keyword Category: Preference	0.236	0.16	0.311	0.22	0.146	0.293	0.203	0.132	0.275	0.154	0.09	0.219	123
Keyword Category: Consideration	0.318	0.179	0.457	0.273	0.14	0.406	0.273	0.14	0.406	0.159	0.05	0.268	44
Keyword Category: This country	0.233	0.159	0.306	0.225	0.152	0.297	0.209	0.139	0.28	0.171	0.105	0.236	129
Giving away stuff	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.667	0.013	1.0	3
Planning (practical questions)	0.429	0.033	0.825	0.429	0.033	0.825	0.429	0.033	0.825	0.429	0.033	0.825	7
Planning to leave after studies/high-school	0.255	0.168	0.342	0.245	0.159	0.33	0.224	0.141	0.308	0.163	0.09	0.237	98
Negative attitude/context	0.256	0.188	0.324	0.231	0.166	0.297	0.225	0.16	0.29	0.2	0.138	0.262	160
General statement (migration related)	0.306	0.266	0.347	0.276	0.236	0.315	0.26	0.221	0.298	0.207	0.171	0.243	493
Indication of a goal (e.g. earn money)	0.417	0.215	0.618	0.375	0.177	0.573	0.333	0.141	0.526	0.333	0.141	0.526	24
Unlikely / obstacles / hypothetical	0.254	0.182	0.325	0.239	0.169	0.31	0.232	0.163	0.302	0.183	0.119	0.247	142
Details provided - work	0.361	0.202	0.52	0.361	0.202	0.52	0.306	0.153	0.458	0.194	0.063	0.326	36
Details provided - education	0.583	0.292	0.875	0.583	0.292	0.875	0.5	0.205	0.795	0.417	0.125	0.708	12
Other details	0.2	0.0	0.41	0.133	0.0	0.311	0.067	0.0	0.197	0.067	0.0	0.197	15
Has friend/family/relative in destination	0.2	0.0	0.592	0.2	0.0	0.592	0.2	0.0	0.592	0.2	0.0	0.592	5
Destination location	0.326	0.189	0.463	0.304	0.17	0.439	0.283	0.151	0.414	0.196	0.08	0.312	46
Time frame indication	0.318	0.205	0.431	0.288	0.178	0.398	0.242	0.138	0.347	0.212	0.113	0.312	66
Multiple tweets	0.383	0.291	0.476	0.327	0.238	0.416	0.318	0.229	0.406	0.271	0.186	0.356	107

A4 Table. Results of logistic regressions for the full and constrained models for 30, 60, 90, and 180 day migration thresholds.

Variable	Migration for at least 30 days	Migration for at least 30 days - restricted	Migration for at least 60 days	Migration for at least 90 days	Migration for at least 180 days	Migration for at least 180 days - restricted
Multiple posts	0.517	0.470	0.350	0.438	0.529	0.524
	0.029	0.043	0.152	0.077	0.043	0.040
keyword category						
Preference	-0.342	-0.476	-0.230	-0.344	0.024	-0.448
	0.400	0.048	0.586	0.419	0.962	0.110
Necessity	-0.266		-0.107	-0.078	0.219	
	0.623		0.846	0.887	0.731	
Planning	1.292	1.128	0.502	0.599	1.444	0.890
	0.056	0.042	0.470	0.389	0.047	0.116
Intention	0.164		0.207	0.072	0.409	
	0.678		0.614	0.862	0.401	
Announcement	0.252		0.222	0.139	0.670	
	0.518		0.584	0.734	0.161	
keyword "this country"	-0.384	-0.339	-0.241	-0.263	-0.303	-0.191
	0.125	0.157	0.341	0.312	0.283	0.470
Context 1	0.159		0.323	0.125	-0.390	
	0.686		0.411	0.760	0.423	
Context 2	1.163	1.029	1.388	1.082	0.861	0.676
	0.080	0.102	0.034	0.094	0.196	0.295
Context 3	-1.063		-1.657	(omitted)	-1.994	-1.506
	0.175		0.115		0.119	0.154
Context 4	-1.240		-1.102	-0.983	-0.567	
	0.294		0.358	0.409	0.633	
Context 5	0.225		0.213	0.149	0.096	
	0.526		0.554	0.688	0.819	
Context 6	-0.389		-0.258	-0.448	-0.322	
	0.228		0.431	0.198	0.373	
Category 1	(omitted)		(omitted)	(omitted)	2.087	
					0.139	
Category 2	0.428		0.701	1.152	1.679	
	0.661		0.481	0.267	0.115	
Category 3	-0.283	-0.405	0.052	0.332	0.218	-0.420
	0.673	0.116	0.941	0.664	0.787	0.161
Category 4	-0.196	-0.214	-0.195	-0.143	0.086	
	0.383	0.328	0.398	0.541	0.729	
Category 5	0.222		0.426	0.734	0.696	
	0.723		0.518	0.311	0.357	
Category 6	-0.065		-0.074	-0.067	0.520	0.431
	0.896		0.885	0.897	0.306	0.373

Variable	Migration for at least 30 days	Migration for at least 30 days - restricted	Migration for at least 60 days	Migration for at least 90 days	Migration for at least 180 days	Migration for at least 180 days - restricted
Category 7	-0.365	-0.350	-0.300	-0.208	-0.115	
	0.120	0.118	0.209	0.390	0.663	
Year 2021 dummy	-0.810	-0.805	-0.610	-0.534	-0.638	-0.591
	0.053	0.050	0.142	0.203	0.177	0.197
Constatnt	-0.928	-0.595	-1.343	-1.652	-2.420	-1.304
	0.199	0.000	0.077	0.043	0.006	0.000
pseudo-R2	0.0445	0.0382	0.0330	0.0267	0.0471	0.0308

First row contains coefficients, second row contains p-values. Values in bold represent variables significant at the 5% level, variables in bold italic represent variables significant at the 10% level.



UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
44/50 DŁUGA ST.
00-241 WARSAW
WWW.WNE.UW.EDU.PL
ISSN 2957-0506