

TNA ACTIVITY REPORT

Dead but Turkish: An NLP Analysis of the Obituaries of the Turkish Community in Belgium

Author: Nisa İrem Kırbaç (nisairemkirbac@gmail.com)

Host institution: Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University

Mentor: Silvie Cinková

Period of stay: 12th June 2023 - 7th July 2023

Introduction

Today, the total number of Turkish people living in Western European countries is approximately 5.5 million (Ministry of Foreign Affairs). According to statistics, the number of Turkish citizens emigrating abroad is increasing every year, reaching the peak of the last 7 years in 2022 (Turkish Statistical Institute, 2023). Although the economic crisis and high employment rates are reported to be the main reasons for the increased population flow—or brain drain—from Turkey to Europe in recent years (Duvar English, 2022), it is crucial to look back to the second half of the twentieth century to explain the high number of people from Turkish origin in Europe today.

The population flow from Turkey to Europe started to dramatically increase in the 1960's, in response to the growing need for labor in the region. Consequently, bilateral agreements, forming the basis for the arrivals of the Turkish population in Belgium, were signed between Turkey and the Western European countries. These initial agreements ensured that the workers' immigrant status in Europe was temporary: they were expected to return home at the end of their contracts, they were not permitted to bring their families with them, and they even needed to live in group living quarters or dormitories allocated to other workers of similar status, rather than in their own homes (Abadan-Unat, 2011, p. 12). By the 1970s, it became clear that the importation of Turkish labor was "no longer a temporary phenomenon" and "the exported labor intended to stay" (Abadan-Unat, 2011, p. 13). As a result, a series of social security agreements ... the foreign workers from *Gastarbeiter* status and granting them social rights including "access to state health care; social insurance in the event of accidents, disability, or death; unemployment and pension rights; and state assistance in childbirth care", were developed and signed (Abadan-Unat, 2011, p. 13). In addition to this shift, the number of children and young migrants in the region has also increased considerably with the introduction of laws encouraging foreign workers to bring their estranged families to join them in Europe. Given this context, a significant part of the Turkish-origin population in Western Europe today is made up of the children, grandchildren and even great-grandchildren of workers who migrated to the region in the 1960s in response to the growing need for labor.



However, despite the fact that Turkish-originated immigrants have been residing in Europe for decades, and they have a very high rate of obtaining citizenship in the European countries they migrate to (Eurostat, 2022), Turkish identity still plays a key role in the way Turkish immigrants define and perceive themselves, and how they interact with their social environments. Christiane Timmerman's ethnographic research among young Turkish women residing in Belgium shows that ethnicity is prominent for Turkish nationals regardless of an individual's level of objective integration in Belgian society and her ideological orientation. Immigrants who are "confronted to a greater extent with 'the other' in society...feel the need to affirm their 'Turkish identity'" (336). Immigrant communities make use of various tools to keep up with *long-distance nationalism* (Anderson, 1998) that "binds together immigrants, their descendants, and those who have remained in their homeland into a single transborder citizenry" (Schiller & Fourn, 2004). In this research we claim that obituaries are one of the tools Turkish migrants use to maintain ties with each other, show their attachment to their roots, and therefore create the Turkish immigrant identity.

Based on the research showing the need among immigrants to build their own community and bond with people with the same cultural values, and obituaries being important in culture, a linguistic analysis of obituaries that belong to the immigrant communities may show us the reflections of specific cultural themes in the context of death. Moreover, analyzing obituaries using NLP tools may enable us to have an understanding of the structure of Turkish obituaries.

Research Questions

Within the context described in the introduction, our initial research questions were as follows:

1. How can we create a corpus of death notices that belong to Turkish-originated migrants in Belgium available in newspapers and the Web (e.g. Twitter)?
2. Kinship and shared values are key themes in the formation and maintenance of a very specific Turkish immigrant identity across generations. What are the expressions used by Turkish-originated immigrants to emphasize their Turkish identity in the obituaries?
3. Although the Turkish nationals have not left Europe after their migration in the 1960s, and even more than three-quarter of them hold citizenship today (Petrovic, 2012), it is possible to say that the generations that continue their lives within the region have not abandoned their national identity. Is it possible to see the reflections of nationalism in the obituaries?

Answering the first of the abovementioned questions was the main goal of my research visit. However, as we modified the duration of my stay in Prague, we also reconsidered the predetermined questions. Instead of trying to analyze the texts with regards to their semantics, we focused on the following questions:

1. How can we create a corpus of death notices that belong to Turkish-originated migrants in *Europe* available in newspapers and the Web (e.g. Twitter)?
2. How can we standardize our corpus to be able to make a meta-data and syntax related comparison between the results obtained from each source?
3. How can we use a language model for Turkish, a lesser sourced language, to conduct NLP analysis on the content of the texts?

Methodology

With the abovementioned questions in mind, our main aim was to create a corpus of death notices that belong to Turkish-originated migrants. This consisted of two steps. The first step was to scrape data from online newspapers that address the population of Turkish origin in

Europe, and the second step was to scrape data from Twitter to obtain a collection of obituaries posted on social media by and/or about immigrants of Turkish origin in Europe. Below we describe the steps used to extract data from two different sources.

Scraping Gazeteturk.be:

- Through a Google search, we identified online newspapers that address Turkish immigrants in Belgium and we came up with a list of five newspapers. The reason behind us specifically focusing on Belgium was because, based on my previous research experience, I am already familiar with the Turkish community in that region. Therefore, initially the aim was to limit the scope of the research to the obituaries that belong to people from Turkish origin in Belgium. Based on the newspaper that we scraped the obituary data from is from one newspaper based in Belgium, Gazeteturk.be.
- Gazeteturk.be, according to their own definition, is an objective and informative online newspaper that serves as a bridge between Europe and Turkey. The reason why we decided to move on with this newspaper was because it has a specific section “Vefat Edenler”, which literally means the “Deceased”, that is dedicated specifically to the death notices. Therefore it was relatively easier to obtain the first part of our corpus basically through scraping the articles on that section.
- To scrape the articles, we used a query based on XPath expressions. As *Image 1* shows this query allowed us to scrape all obituaries in the section. In addition to the main content, we obtained metadata containing the language, title, publishing time, author, and tags of the posts.

```

<newspaper>
<title ident="gazeteturk">Gazete Türk</title>
<language ident="tur">Turkish</language>
<country ident="TUR">Turkey</country>
<baseurl>https://gazeteturk.be/</baseurl>
<content>
  <title>//meta[@property="og:title"]/@content</title>
  <date>//meta[@property="og:updated_time"]/@content</date>
  <author>//meta[@property="article:author"]/@content</author>
  <section>//div[@class="safirCategoryMeta"]/a</section>
  <text>//*[@id="singleContent"]/p</text>
  <tags>//div[@class="tags"]</tags>
  <desc>//meta[@name="description"]/@content</desc>
  <exclude>//script</exclude>
  <exclude>//figure</exclude>
  <exclude>//iframe</exclude>
  <exclude>//img</exclude>
</content>
<crawl agent="Mozilla/5.0">
  <startpage>https://gazeteturk.be/category/vefatedenler/</startpage>
  <item>//div[@class="postthumb"]/a/@href</item>
</crawl>
</newspaper>

```

Image 1: XPath query for the articles in the “Vefat Edenler” section of Gazeteturk.be

- In total we obtained 179 obituaries published on Gazeteturk.be in total: the oldest of the articles is from June 1st, 2021 and the most recent one is from June 14th, 2023. Using Text Encoding Initiative (TEI) guidelines to maintain a standard, we uploaded the texts to TEITOK, an online platform for allowing the collection, search and analysis of corpora currently maintained at ÚFAL (Charles University).

Scraping Twitter

- As the first step of the study, a manual search was carried out on Twitter to have a more concrete understanding of the general features of obituary tweets. After using more general queries such as *ölüm ilanı* (death announcement), *vefat* (death), *kaybettik* (we lost), we moved on to more specific queries. Based on our initial goal of focusing only on Belgium, we modified the queries and included *Belçika* (Belgium). However using the combination of previous words with *Belçika* (i.e. *Belçika death*) led us to posts of newspaper articles about any death in Belgium instead of responding to our research question, “Which specific search queries can allow us to have a collection of obituaries posted by or about Turkish origin immigrants in Belgium?”. Therefore, we decided to use more specific and traditional words that are used in Turkish language and that are only used when referring to the death of a Muslim person. When we modified the search by including the words *rahmet* (mercy) and *mekanı* (residence with possessive suffix)¹ in the keywords, we achieved the closest results to our goal. We made the decision to use two specific query combinations to create the second part of our corpus: *Belçika vefat mekanı* and *Belçika vefat rahmet*.
- Following the decision on the queries to use, we moved on to our next step: finding the best tool to scrape Twitter. Scraping Twitter was not as easy as scraping the newspaper, mostly due to the most recent regulations on the website regarding very limited free API access. As we could not find a proper Python package to do obtain the data, we decided to subscribe for the basic tier of API that “offers low level of API usage, and access to Ads API for a \$100 monthly fee” (XDevelopers, 2023). Unfortunately, as the data that we could scrape via this subscription was limited to the tweets from the past seven days, we decided to try finding another way.
- sncrape: After a long search, we came up with sncrape, a scraper that “scrapes things like user profiles, hashtags, or searches and returns the discovered items” (sncrape, 2023). For Twitter, the information returned contains “users, user profiles, hashtags, searches (live tweets, top tweets, and users), tweets (single or surrounding thread), list posts, communities, and trends” (sncrape, 2023). Using sncrape we ran our commands directly in the terminal, and ended up scraping the data for the two search queries into two separate JSON files.
- Change of plans: One problem that we realized at this point was the amount of tweets was very limited (45 tweets for *Belçika vefat rahmet* and 12 tweets for *Belçika vefat mekanı*). At this point, we were already thinking about extending the scope of the research and having a corpus of obituaries of Turkish origin people in Europe, instead of just Belgium. We were also well aware that we will not have enough time to analyze the data. So, we decided to modify our goal at this stage of the project to create a strong corpus and make the corpus analyzable with the help of UD Pipe, instead of analyzing the results as originally thought.
- To this end, we revised our queries to include the three other countries with the highest number of Turkish origin immigrants (Germany, France, and the Netherlands). In addition, we also searched for the word *gurbetçi* (someone with a Turkish origin who lives and works in another country, especially in Germany) to obtain immigrant-oriented results rather than region-oriented results. In this way, in addition to the 57 tweets we obtained as a result of the searches for *Belçika vefat mekanı* (12 tweets) and *Belçika vefat rahmet* (45 tweets), we obtained a total collection of 1860 tweets by running our command for *Almanya vefat mekanı* (1063 tweets), *Almanya vefat rahmet* (392 tweets), *Fransa vefat mekanı* (21 tweets), *Fransa vefat rahmet* (141 tweets), *Fransa vefat mekanı* (16 tweets),

¹ “Allah rahmet eylesin” meaning “May Allah be merciful”, and “Mekanı cennet olsun” meaning “May they reside in Heaven” are two expressions used in Turkish to express sorrow for the death of specifically Muslim people.

Fransa vefat rahmet (82 tweets), gurbetçi vefat rahmet (12 tweets) and gurbetçi vefat rahmet (76 tweets).

- Next, we converted JSON files into meaning making tables using R. We ended up having six tables per query: MainTweetContent, MainTweetAuthor, MainTweetHashtags, GPSPMainTweet, LinksFromMainTweet and inReplyToUser. *Image 2* is an example of these tables, only containing the information that we need for our research.

tweet_id	renderedContent	url	date
100186420108552400	Beklenenin aksine Arınca'nın 2020 yılını bile geçilmeden yaşamını yitirdi. Bu ölümler omuzlarında yıkılmadı. Allah onları rahmetine kavuşturulsun.	https://twitter.com/ArincaninMevlasi/status/100186420108552400	2020-05-25T09:29:40+00:00
100229790100326200	Yusuf Merali'nin ölümü Fransa'da bir vefat değil, Türkiye'de bir vefat. Çünkü Yusuf Merali, Türkiye'nin en büyük şairlerinden biri. Onun ölümü, Türkiye'nin büyük bir kaybıdır.	https://twitter.com/yusufmerali/status/100229790100326200	2020-11-14T09:30:00+00:00
100346407000000000	Fransa'da hayatını kaybeden şair Yusuf Merali'nin ölümü, Türkiye için büyük bir vefat. Çünkü Yusuf Merali, Türkiye'nin en büyük şairlerinden biri.	https://twitter.com/yusufmerali/status/100346407000000000	2020-11-29T10:20:00+00:00
100346407000000000	Fransa'nın ölümlerini takip eden günün 23.10.2020 tarihinde ölen Yusuf Merali'nin ölümü, Türkiye için büyük bir vefat. Çünkü Yusuf Merali, Türkiye'nin en büyük şairlerinden biri.	https://twitter.com/yusufmerali/status/100346407000000000	2020-10-29T13:20:00+00:00
107224447000000000	Gurbetçi Çerçen Sakarya Terleştirdiği Ölümler Arınca Çerçen'in 4 gün sonra yaza kurbanı olarak hayatını yitirdi. Haberi deniz bir gazeteye öğrendim. Merhumu Allah'tan rahmet, ailesine ve yakınlarına başsağlığı dilerim.	https://twitter.com/merhumuAllah/status/107224447000000000	2020-09-20T16:00:40+00:00
152003004254419	Arınca'nın ölümü büyük bir vefat. Çünkü Arınca, Türkiye'nin en büyük şairlerinden biri. Onun ölümü, Türkiye'nin büyük bir kaybıdır.	https://twitter.com/ArincaninMevlasi/status/152003004254419	2020-09-20T14:41:17+00:00
14819604902970718	Onun ölümü büyük bir vefat. Çünkü Arınca, Türkiye'nin en büyük şairlerinden biri. Onun ölümü, Türkiye'nin büyük bir kaybıdır.	https://twitter.com/ArincaninMevlasi/status/14819604902970718	2020-09-20T10:20:00+00:00
144079070000000000	Gurbetçi Çerçen Sakarya Terleştirdiği Ölümler Arınca Çerçen'in 4 gün sonra yaza kurbanı olarak hayatını yitirdi. Haberi deniz bir gazeteye öğrendim. Merhumu Allah'tan rahmet, ailesine ve yakınlarına başsağlığı dilerim.	https://twitter.com/merhumuAllah/status/144079070000000000	2021-10-29T13:20:00+00:00
142732410000000000	Arınca'nın ölümü büyük bir vefat. Çünkü Arınca, Türkiye'nin en büyük şairlerinden biri. Onun ölümü, Türkiye'nin büyük bir kaybıdır.	https://twitter.com/ArincaninMevlasi/status/142732410000000000	2021-08-18T17:44:30+00:00
142231437100000000	Gurbetçi Çerçen Sakarya Terleştirdiği Ölümler Arınca Çerçen'in 4 gün sonra yaza kurbanı olarak hayatını yitirdi. Haberi deniz bir gazeteye öğrendim. Merhumu Allah'tan rahmet, ailesine ve yakınlarına başsağlığı dilerim.	https://twitter.com/merhumuAllah/status/142231437100000000	2021-08-02T15:11:21+00:00
142231437100000000	Gurbetçi Çerçen Sakarya Terleştirdiği Ölümler Arınca Çerçen'in 4 gün sonra yaza kurbanı olarak hayatını yitirdi. Haberi deniz bir gazeteye öğrendim. Merhumu Allah'tan rahmet, ailesine ve yakınlarına başsağlığı dilerim.	https://twitter.com/merhumuAllah/status/142231437100000000	2021-08-02T15:11:21+00:00
142231437100000000	Gurbetçi Çerçen Sakarya Terleştirdiği Ölümler Arınca Çerçen'in 4 gün sonra yaza kurbanı olarak hayatını yitirdi. Haberi deniz bir gazeteye öğrendim. Merhumu Allah'tan rahmet, ailesine ve yakınlarına başsağlığı dilerim.	https://twitter.com/merhumuAllah/status/142231437100000000	2021-08-02T15:11:21+00:00
142231437100000000	Gurbetçi Çerçen Sakarya Terleştirdiği Ölümler Arınca Çerçen'in 4 gün sonra yaza kurbanı olarak hayatını yitirdi. Haberi deniz bir gazeteye öğrendim. Merhumu Allah'tan rahmet, ailesine ve yakınlarına başsağlığı dilerim.	https://twitter.com/merhumuAllah/status/142231437100000000	2021-08-02T15:11:21+00:00

Image 2: MainTweetContent table for the results of gurbetçi vefat rahmet query

```

1 import pandas as pd
2 from lxml.etree.ElementTree import Element, SubElement, ElementTree
3
4 tv_files = ['Users/nisaireskirbac/Desktop/sandbox/results_validjson_gurbetci_vefat_rahmet/MainTweetAuthor-validj
5 data_dict = {}
6 dtype_dict = {
7     'tweet_id': object,
8     'usr_followersCount': int,
9     'usr_friendsCount': int
10 }
11
12 for file in tv_files:
13     data = pd.read_csv(file, delimiter=';', dtype=dtype_dict)
14     tweet_id_column = [col for col in data.columns if 'tweet_id' in col.lower()][0]
15     for index, row in data.iterrows():
16         tweet_id = str(row[tweet_id_column])
17         if tweet_id not in data_dict:
18             data_dict[tweet_id] = {}
19             data_dict[tweet_id][file] = row
20
21 for tweet_id, tweet_data in data_dict.items():
22     tei = Element('TEI')
23     teiHeader = SubElement(tei, 'teiHeader')
24     fileDesc = SubElement(teiHeader, 'fileDesc')
25     titleStm = SubElement(fileDesc, 'titleStm')
26     sourceDesc = SubElement(fileDesc, 'sourceDesc')
27     bibl = SubElement(sourceDesc, 'bibl')
28     profileDesc = SubElement(teiHeader, 'profileDesc')
29     text = SubElement(tei, 'text')
30
31 for file, data in tweet_data.items():
32     if 'MainTweetAuthor' in file:
33         title = SubElement(titleStm, 'title')
34         title.text = str(data['tweet_id'])
35         author = SubElement(titleStm, 'author')
36         author.text = str(data['username'])
37
38     for file, data in tweet_data.items():
39         if 'MainTweetContent' in file:
40             idno_tweetID = SubElement(bibl, 'idno', type='tweetID')
41             idno_tweetID.text = str(data['tweet_id'])
42             date = SubElement(bibl, 'date')
43             date.text = str(data['date'])
44
45             elif 'MainTweetAuthor' in file:
46                 ident_usr_location = SubElement(bibl, 'country', ident='ISO')
47                 ident_usr_location.text = str(data['usr_location'])
48                 author = SubElement(bibl, 'author')
49                 name = SubElement(author, 'name')
50                 name.text = str(data['username'])
51                 idno_userid = SubElement(author, 'idno', type='userid')
52                 idno_userid.text = str(data['usr_id'])
53                 idno_displayname = SubElement(author, 'idno', type='displayname')
54                 idno_displayname.text = str(data['usr_displayname'])
55                 note_followers = SubElement(author, 'note', n='followers')
56                 note_followers.text = str(data['usr_followersCount'])
57                 note_following = SubElement(author, 'note', n='following')
58                 note_following.text = str(data['usr_friendsCount'])
59
60 for file, data in tweet_data.items():
61     if 'MainTweetContent' in file:
62         sourceurl = SubElement(profileDesc, 'sourceurl')
63         sourceurl.text = str(data['url'])
64         langUsage = SubElement(profileDesc, 'langUsage')
65         ident_language = SubElement(langUsage, 'language', ident='tur')
66         ident_language.text = str(data.get('Turkish', 'Turkish'))
67
68 for file, data in tweet_data.items():
69     if 'MainTweetContent' in file:
70         renderedContent = SubElement(text, 'text')
71         renderedContent.text = str(data['renderedContent'])
72
73 tree = ElementTree(tei)
74
75 output_file = f'tweet_{tweet_id}.xml'
76
77 tree.write(output_file, encoding='utf-8', xml_declaration=True)
78
79
80
81
82

```

Image 3: Code to transform the information in the JSON files to TEITOK format

- As the next step, we uploaded our results to TEITOK. However, since the data we obtained with the help of snscraper was in JSON format, and we converted them into tables, we still had to extract the necessary information from each table in accordance with TEI standards. So, we wrote and ran a code allowing us to make each tweet suitable for uploading to TEITOK. This code can be seen in *Image 3*.

Description & Outcomes of the Research Visit

Weeks / Days	Week 1	Week 2	Week 3	Week 4
1	Deciding on the plan	Optimizing the queries that can be used to get the best result on Twitter	Scraping Twitter using snsrape	Uploading data to TEITOK
2	Scraping Gazeteturk.be		Converting JSON files to tables	Using NLP tools to tokenize, lemmatize, and tag the corpus. Working on various NER models for Turkish language
3		Trying multiple tools to scrape Twitter		Trying out some interactive queries in Corpus Query Language to extract specific information from the obituaries
4	Workshop on Data Analysis		Converting the information in the tables into TEI format	Considering future work
5		Scraping Twitter using snsrape		

Table 1: Table showing weekly activities

As indicated before, our original plan was that my visit to Charles University would last 3 months. However, due to some bureaucratic problems, we had to modify this plan and reduce the duration of the visit to one-month. Therefore, the research process and the evaluation of the data, which was initially intended to be much more detailed and inclusive, had to be narrowed down to fit the duration of the visit.

The 1-month plan in line with the objectives of the study can be seen in *Table 1*. The first three weeks of the visit were dedicated to extracting data from two different sources and building the corpus. Meanwhile, the aim was to answer the first two research questions indicated above. The steps followed to achieve this goal are described in detail in the methodology section of this report. I also attended a three-day workshop on Data Analysis organized by Charles University in the first week. During the workshop, I had the opportunity to learn about the ongoing research in the field of text analysis carried out by the Faculty of Mathematics and Physics and thus to reflect on the applications of computational tools, and specific NLP tools, in the social sciences and humanities.

Following the acquisition of the corpus, the last week of the visit was dedicated to preparing the texts for analysis with the help of NLP tools. In this context, tokenization, lemmatization and tagging using Universal Dependencies were performed via TEITOK (see *Image 4*). Moreover, several NLP models for the Turkish language were tested to find the model that best recognizes elements such as names, locations, numbers and dates in our data. Once all the files have been parsed and the corpus has been indexed to be available to search, we worked on the possible interactive queries in Corpus Query Language (see *Image 5*). This helped us to have an initial understanding of how obituaries that belong to Turkish immigrants are structured in terms of the information that they contain, and if there are some patterns regarding the relative positioning of this information.

CQL Query: Search query builder | visualize | options

154 results • ipm: 2905.61 • Showing 0 - 100 • next

Tags:

context Müftü oğullan) oğlu, 1927 doğumlu 95 yaşındaki Türkiye'nin ilk onkoloji (Kanser
context ve Topcuoğlu İbrahim Yedikapu kızı 1927 doğumlu 95 yaşındaki Sultan Yedikapu hakkın rahmetine kavuşmuştur
context (Topuz) eşi, 1928 doğumlu 93 yaşındaki Ceylan Korkmaz hakın rahmetine kavuşmuştur
context ve Bayram Çeliker oğlu, 1929 doğumlu 93 yaşındaki Halil Çeliker hakkın rahmetine kavuşmuştur
context Afyonkarahisar'ın Emirdağ ilçesi nüfusuna kayıtlı 1930 doğumlu 91 yaşındaki Kamil Ün isimli Türk vatandaşı
context ve Abdullah Kadir kızı, 1930 doğumlu 92 yaşındaki Şükriye Akbalı (Imre)
context ve Salih Bostan oğlu, 1930 doğumlu 92 yaşındaki Talat Bostan hakkın rahmetine kavuşmuştur
context ve Mevlüt Özgüler kızı, 1930 doğumlu 93 yaşındaki Gülsüm Tapmaz(Özgüler) hakkın rahmetine
context Abdil Kadir Çivrilili eşi, 1931 doğumlu 90 yaşındaki Dudu Çivrilili hakkın rahmetine kavuşmuştur
context , Güzel ve Mukaddin kızı 1931 doğumlu 90 yaşındaki Mecbure Yüksel hakkın rahmetine kavuşmuştur
context Emirdağ esnafı Potuklardan Mehmet Çil 1931 doğumlu 91 yaşındaki Emirdağ esnafı Potuklardan mehmet Çil
context Fatma(göğücü) ve Kamil kızı 1931 doğumlu 91 yaşındaki Şükriye Halaç hakkın rahmetine kavuşmuştur
context Fatma ve Mehmet kızı, 1932 doğumlu 90 yaşındaki Arzı Özkaya hakkın rahmetine kavuşmuştur
context , Zehra Köroğlu eşi, 1933 doğumlu 88 yaşındaki Dilaver Köroğlu hakkın rahmetine kavuşmuştur
context ve İsa Yürük oğlu, 1933 doğumlu 88 yaşındaki Süleyman Yürük (Sülo Amca
context Zakire ve Süleyman Öztürk oğlu 1933 doğumlu 88 yaşındaki Abdülkadir Öztürk hakkın rahmetine kavuştu

Image 6: Result of the query written to extract Birth of Date and Age of Death information for Gazeteturk.be articles

In conclusion, after a month of research we have come a long way towards creating a corpus of obituaries of people of Turkish origin who have migrated to Europe. In this process, we have realized that the texts to be analyzed should come from at least two different sources. Based on our initial observations, we can say that the most important reason for this is the genre-based differences between the obituaries published in newspapers and the ones that can be found on social media sites. According to our initial observations, obituaries published in newspapers, where the content is more controlled, are more in line with Bawarshi's definition of obituary; "... notices of a person's death, usually accompanied by a short biographical account" mainly aim to notify the public (Bawarshi, 2000). On the other hand, those published on social media are closer to the definition of eulogy, aiming "to assess and praise the meaning of the deceased's life and death" (Bawarshi, 2000). Therefore, we need to take this into account for our future work.

Limitations & Considerations over Future Work

Limitations

The three main limitations of our research are:

- Obituaries that we scraped from Gazeteturk.be are very limited. This limitation is two fold. Firstly, although some of the obituaries are about Turkish immigrants who passed away in different European countries, Gazeteturk.be is a Belgium-based publication. As we would like to focus on the whole of Europe, instead of focusing only on Belgium, this is a crucial problem. Secondly, as the website lost their archive for the articles before 2021, the data that we obtained is very limited in terms of number. Detecting, obtaining and analyzing the death notices on Gazeteturk.be was less complicated than doing so for the Twitter posts. Therefore we need to find more inclusive newspapers and try to have more data.

- We ended up applying an NER model to the corpus. However, the model is still very far from being perfect. Therefore the need to focus on testing new models, or bettering the model that we already have is very crucial.
- And thirdly, although we have already started working on the syntactic analysis of the corpus, and ended up having some fairly good results for the data obtained from Gazeteturk.be, the data obtained via Twitter is very much different. As there are almost no rules with regards to the structure of the Twitter posts, and there is way more to discover (i.e. emojis, hashtags), we need to spend more time on understanding the tweets.

Considerations over Future Work

In addition to extending our corpus, some of the areas we plan to address in our future work are:

- To examine the NER models developed for Turkish in more detail and to provide suggestions for the improvement of these models
- In the light of the research questions we identified prior to the reduction of the duration of the visit, to analyze our corpus in terms of a) biographical and practical information covered in the texts, and b) reflections on immigrant culture, nationalism, and solidarity.

Acknowledgements

First of all, I would like to express my special gratitude to my mentors at Charles University, Prof. Silvie Cinková, for her guidance, support, but most of all her generous encouragement throughout the process. I would like to thank Maarten Janssen for his technical support and patience. I would like to thank the entire Institute of Formal and Applied Linguistics (ÚFAL) team for welcoming me.

Next, I want to thank the CLS INFRA team for allowing me to meet very valuable academics and making this visit possible.

Finally, I want to thank Prof. Julie Birkholz for making me aware of this program and for her support during my application process.

References

- Abadan-Unat, N. (2011) *Turks in Europe: From guest worker to transnational citizen*, Translated by C. Campion. New York: Berghahn Books.
- Anderson, B. (1998) 'Long-Distance Nationalism', in *The spectre of comparisons : nationalism, Southeast Asia, and the world*. Verso, pp. 58–76.
- Bawarshi, A. (2000) 'The Genre Function', in *College English*. (62:3), pp. 335–360.
- Duvar English* (2022) 'Report finds 73 percent of young people in Turkey want to live abroad', 16 February. Available at: <https://www.duvarenglish.com/report-finds-73-percent-of-young-people-in-turkey-want-to-live-abroad-news-60382> (Accessed: 16 August 2023).
- Eurostat (2022) *Enlargement countries - statistical overview*. (Accessed: 14 August 2023).
- JustAnotherArchivist (2023) *snsrape*, *github*. Available at: <https://github.com/JustAnotherArchivist/snsrape> (Accessed: 17 August 2023).
- Petrovic, M. (2012) 'Belgium: A Country Of Permanent Immigration', *The Online Journal Of The Migration Policy Institute* [Preprint]. doi:<https://www.migrationpolicy.org/article/belgium-country-permanent-immigration>.
- Republic of Turkey Ministry of Foreign Affairs. *Turkish Citizens Living Abroad*. Available at <http://www.mfa.gov.tr/the-expatriate-turkish-citizens.en.mfa>. Accessed 20 May 2022.
- Schiller, N.G. and Fourn, G.E. (2001) *Georges woke up laughing: Long-distance nationalism and the search for home*. Durham, NC: Duke Univ. Press.
- Timmerman, C. (2000) 'Secular and religious nationalism among young Turkish women in Belgium: Education may make the difference', *Anthropology & Education Quarterly*, 31(3), pp. 333–354. doi:10.1525/aeq.2000.31.3.333.
- Turkish Statistical Institute (2023) *Immigrants and emigrants by citizenship and provinces, 2016-2022*, TURKSTAT. (Accessed: 15 August 2023).
- XDevelopers (2023) [Twitter] 9 February. Available at: https://twitter.com/XDevelopers/status/1623467615539859456?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1623467617037221889%7Ctwgr%5E011afa6df4157c50d9824aa2c15218d0d3810a0b%7Ctwcon%5Es2 &ref_url=https%3A%2F%2Ftechcrunch.com%2F2023%2F02%2F08%2Ftwitter-says-the-basic-tier-of-its-api-will-cost-100-per-month%2F (Accessed: 17 August 2023).