Contact-Induced Linguistic Change in EU Polish:
A Corpus Study of Legal, Institutional and Political Discourse

Aleksandra Tomaszewska

University of Warsaw

July 15, 2024

The European Union and Languages

- The EU unifies 27 nations with 24 official languages.
- Institutional multilingualism as an emblem of the EU (e.g., TEU)
- Responses in citizens' chosen EU language.
- EU communication spans various genres, subjects, and domains.
- Major working languages: French, German, and English.

Role of translators/interpreters

- Translation/interpretation – crucial component of communication
- Genres – translator-mediated
- Influence on national languages within genres

EU's Linguistic Landscape

- Interlingual contact
- Distinct linguistic style within the EU.
- 'Eurofog' 'Eurobabble' 'Eurojargon' 'Eurolect' (Goffin 1990: 18)
- Changes in national languages (EU-induced linguistic change; effect e.g., EU borrowings)

Rationale

- Criticized for vagueness and complexity. Prescriptive observations.
- Rationale: Limited empirical; non-legal; multi-genre; corpus data on EU-induced linguistic change.

Research Focus

- Getting new knowledge on EU Polish (trends).

Key Research Questions

- What are the effects of EU-induced language change in the corpus of EU law, press releases, and interviews with Members of the European Parliament in Polish, and what is their distribution in the subcorpora?
- Do they integrate into Polish and, if so, what trends can be discerned?
- What are their linguistic features and functions in contexts?
- How has their usage evolved within Polish over time, as analyzed through microdiachronic and diachronic data?

Corpus Data

- Research Material:

- I created a monolingual "EU Polish" corpus with 8 mln words and 2500 texts (5 years).

- EU law, institutional press releases, and MEP interviews.

- 2 corpora of translations (specialized and non-specialized).

Data Sources:

- Directives and regulations from Eur-LEX

- Press releases from EC archives.

- Transcriptions of TV interviews with MEPs on popular programs.
    - Additional data used:
    - ParlaMint Open Access Corpus
    - National Corpus of PL and pl-Ten-Ten as reference
    - EU law corpora until 2016 available in Sketch Enigine

Identification

- Methods Used:
  - Semi-automatic and manual verification.
  - Sketch Engine.
  - Wordlist analysis, including hapax legomena.
  - Term extraction against a reference corpus.
  - Searching for legislative definitions.
  - Manual analysis to identify foreign elements.
  - Extraction of acronyms and their analysis.
- Challenges:
  - No effective software for automatic extraction. (General problem in language contact studies, also in Witalisz and Mańczak-Wohlfeld)

Results: Distributional Analysis

- Distribution of EU Borrowings:
  - Total EU borrowings identified: 474 (numbers and percentage for approximation only)
    - EU law: 246 (0.3%)
    - Press releases: 175 (0.18%)
    - Interviews: 53 (0.07%)
- Trends in Distribution:
  - Limited presence of EU borrowings in Polish.
  - Highest presence in EU law, followed by press releases.
  - Lowest presence in TV interviews with MEPs.
  - Nonce borrowings (pl. okazjonalizmy).
  - Co-existing variants.
  - Often signaled in contexts.
  - Institutionalized calques; some in the IATE database.

Results: Integration

- Aim: Assess the degree of integration of the effect of EU-induced linguistic change in PL.
- Classification of Integration Levels:
  - Pronunciation
  - Spelling
  - Morphology (inflection, word-formation)
  - Meaning
- Levels of Integration:
  - None
  - Partial
  - Full
- Points Assigned:
  - 0 points for non-integrated.
  - 1 point for partially integrated.
  - 2 points for fully integrated.

Results: Integration

- Total results of integration:
    - 60 per cent fully integrated
    - 22 per cent not integrated
    - 18 per cent partially integrated
- Limited impact on foreignness of Polish
- Trends in distribution
- Potential interpretation difficulties
- Highlighting the challenges of linguistic interpretation in the context of EU Polish.

Results: Integration of EU Borrowings

- co-badging – when more than one payment brand is assigned to a payment instrument (Law)
- bio-prospecting - biodiversity prospecting or bioprospecting is the systematic search for biochemical and genetic information in nature in order to develop commercially-valuable products for pharmaceutical, agricultural, cosmetic and other applications. (Press)
- eurobonds (Interviews) - A Eurobond is a debt instrument that's denominated in a currency other than the home currency of the country or market in which it is issued.

Assimilated units:

- zielona ksiega/green paper – Green papers are documents published by the European Commission to stimulate discussion on given topics at European Union (EU) level (Press)
- substandard (Interviews, partially integrated)
- cienki klient [thin client] - a thin client is a simple (low-performance) computer that has been optimized for establishing a remote connection with a server-based computing environment

Results: Use/Functions of EU Borrowings

- Functions related to contact-induced language change
- Examples: lexical gap-filling, prestige, cognitive proximity, etc.
- Genre-related functions
- Ad hoc and code-switching uses

Results: Use of EU Borrowings

- Consideration in institutional terms.
- In interviews, not always derived from institutional guidelines due to genre features.
- Some uses dictated by institutional recommendations, such as presence in IATE.
- EU borrowings serve various purposes:
  - Ensuring the uniform authenticity of a legal act.
  - Filling lexical gaps when native words don't exist.
  - In nearly 30 per cent of cases, the function is to fill lexical gaps, the most frequent function.

Usage Analysis – Procedure using WordSketch

- Detailed linguistic analysis of EU borrowings.
- Used a multi-step procedure for analysis.
- Word sketches generated for each item using Sketch Engine.
- Collocates and grammatical relations extracted.
- Results verified in two stages to minimize bias.
- Typicality considered for comparing genres.
- WordSketch data provided insights into linguistic behavior.
- Revealed syntactic calques and unidentified collocations.

Agenda Usage Analysis

- The term "agenda" is frequently used in institutional and political contexts.
- Common collocations include "cyfrowa" (digital agenda), "Europejska" (European agenda), "2030" (2030 agenda), "poczynić" (to make an agenda), and "polityczny" (political agenda).
- "Cyfrowa" (digital agenda) shows a semantic extension, implying a broader sense of goals and objectives.
- "Agenda" often refers to planning and setting goals in various initiatives.
- Strong semantic connections with EU ideas and initiatives are evident.

EU-induced linguistic change: color calques

- People associate colors with emotions, ideas, and concepts, which may vary across cultures.
- Linguistic economy is one reason for encoding additional meanings in color names and derived adjectives.
- European languages have adopted figurative meanings from color terms
- The European Union's linguistic homogeneity contributes to uniform color term interpretation.
- Two-word borrowings combining color adjectives and nouns were examined across genres.

Frequent Color Term Calques

- 36 different calques with polysemous color-related adjectives were identified.
- Here are some other examples:

Color calques: Blue

- **Literal Meaning:** The color blue.
- **Figurative Meanings:**
  - Primarily used in the context of maritime economy.
  - Example: "niebieska ksiega" (the blue book).
  - Represents sea-related activities.

Color calques: Green

- **Literal Meaning:** The color green.
- **Figurative Meanings:**
    - Associated with ecology and environmentally friendly concepts.
    - Example: "zielona energia" (green energy).
    - Symbolizes positivity and sustainability.
    - zielona ksiega (green book) - 65 in press releases, 7 in EU law
    - zielona infrastruktura (green infrastructure) - 9 in press releases
    - zielona stolica (green capital) - 31 in press releases

Aim of Diachronic Study

- Investigate the diachronic evolution of EU borrowings in EU Polish.
- Analyze EUR-Lex corpus with timestamps.
- Two research phases:
  - Large EUR-Lex corpus analysis.
  - Microdiachronic corpus (2005, 2013, 2021) for deeper insights.
- Examine language changes alongside socio-political developments.

Prefixes - Upward Trend

- EU borrowings beginning with prefixes, including international prefixes, have experienced a significant increase over the years.

- Examples of EU borrowings with prefixes showing a sharp upward trend:

EU Borrowing	Translation	Trend	Significance
biopaliwa	'biofuels'	3.08	1211 (sharp increase)
transgraniczna	'cross-border'	2.90	990 (sharp increase)
mikroprzedsiebiorstw	'micro-enterprises'	2.90	1458 (sharp increase)

Calques - Upward Trend

- Several EU calques have seen a sharp increase in use.
- Examples with the highest trend include:

Calque	Translation	Trend	Significance
incydentalny	'incidental'	3.270852618	2691 (sharp increase)
innowacyjne	'innovative'	2.904210878	3471 (sharp increase)
transponowane	'transposed'	2.904210878	455 (sharp increase)
monitorujaca	'monitoring'	2.904210878	267 (sharp increase)
konsolidowane	'consolidated'	2.904210878	168 (sharp increase)

Evolution: Word Usage Trends

- **Positive Usage Trends:**
  - Upward; EU borrowings with international prefixes.
  - Emergence of new phenomena like cross-border cooperation and green initiatives.
  - Rise in the use of calques related to legal issues, innovation, consolidation, and incidentality.
  - No increase in unadapted foreign words in EU law.

- **Negative Usage Trends:**
  - Decrease in redundant words accompanying legal terms (plain languge promotion effect?).
  - Significant reduction in unadapted borrowings (institutionalization of translation?).

Results: Example - Europejski Semestr

- Pseudo-anglicism from EU jargon.
- Denotes the coordination of economic, fiscal, employment, and social policies in the EU.
- Most important part occurs in the first 6 months of the year.
- Functions in this context: economy, filling lexical gaps, prestige.

Results: EU Borrowings in MEP Samples

- Sample 16:
  - "Ale ludzie musza poczuć, że jest jakiś window of opportunity..."
  - Translation: "But people must feel that there's some kind of a window of opportunity..."
- Sample 18:
  - (Sample 18) Nie ma sensu eksperymentować z kimś nieznanym, no name'em, badź z sfrustrowanym outsiderem, który bedzie chciał na chwile wrócić.
  - It does not make sense to experiment with someone unknown; with a no name or with a frustrated outsider who will want to come back for a moment] (Jacek Kurski, Kontrwywiad, RMF FM, October 1st, 2012)
- (Jacek Kurski, Kontrwywiad, RMF FM, October 1st, 2012):
  - "W transpozycji i implementacji acquis communautaire mamy mixed record"
  - Translation: "As for the transposition and implementation of the 'acquis communautaire', we have a mixed record"
- "twarde jadro ["the hard core of the EU"],"
- "I dlatego pozostaje wrażenie, że Polska jest w narożniku..."

Conclusions
Key Findings

- EU-induced linguistic change in EU Polish genres was examined in the corpus of EU law, press releases, and interviews with MEPs.
- Four main aspects were analyzed: distribution, integration, functions, and evolution of linguistic changes.
- These findings contribute to a better understanding of language dynamics in multilingual settings (+ not only focused on law).

Conclusions
Overall Insights

- EU-induced linguistic change is noticeable in institutional genres but not causing significant hybridization.
- Not so many EU borrowings in the corpora. 60 per cent of lexical items were fully assimilated, indicating a high level of integration.
- Occurrence of EU borrowings in oral speech is infrequent and linked to ad hoc needs.
- Translators play a crucial role as 'linguistic filters' and EU borrowings serve diverse functions across genres, mainly regulated by institutional practices.
- EU borrowings have evolved over time, with some older terms (rather than non-terms) still in use.
- Newer, less-jargon forms have replaced some older borrowings.

Future Research

- **Continuing Research on Common Polish:**
  - Explore EU-ILC in non-institutional contexts.
- **Big Data/NLP:**
  - Utilize NLP for broader insights on more extensive material.
- **Internal Institutional Research:**
  - Collaborate with institutions.
  - Partner with translators/terminologists for linguistic expertise.

References

1. Bańko, Mirosław, Diana Svobodová, Joanna Raczaszek-Leonardi, and Marcin Tatjewski. 2016. Nie Całkiem Obce. Zapożyczenia Wyrazowe w Jezyku Czeskim i Polskim. Warsaw: WUW.

2. Cierpich, Agnieszka. 2019a. Zapożyczenia angielskie w polszczyźnie korporacyjnej. Wydawnictwo Naukowe Akademium Ignatianum.

3. Kjær, Anne Lise. 2007. "Legal Translation in the European Union: A Research Field in Need of a New Approach." In Language and the Law: International Outlooks, edited by Krzysztof Kredens and Stanisław Goźdź-Roszkowski, 69–95. Frankfurt am Main: Peter Lang.

4. Mańczak-Wohlfeld, Elżbieta. 2006. Angielsko-polskie kontakty jezykowe. Wydawnictwo UJ.

5. Matras, Yaron. 2009. Language Contact. Cambridge University Press.

6. Mori, Laura. 2018. Observing Eurolects: Corpus Analysis of Linguistic Variation in EU Law. Studies in Corpus Linguistics 86. Amsterdam, Netherlands: John Benjamins Publishing.

7. Witalisz, Alicja. 2016. Przewodnik Po Anglicyzmach w Jezyku Polskim. Towarzystwo Miłośników Jezyka Polskiego.

Thank You & Q&A Session

Thank you for your attention. Questions?
Contact: a.tomaszewska2@uw.edu.pl