

# A pilot study: Application of Market Basket Analysis to Text Analysis

Václav Cvrček, Charles University

Masako Fidler, Brown University

Threat-Defuser Plenum meeting

March 9, 2021

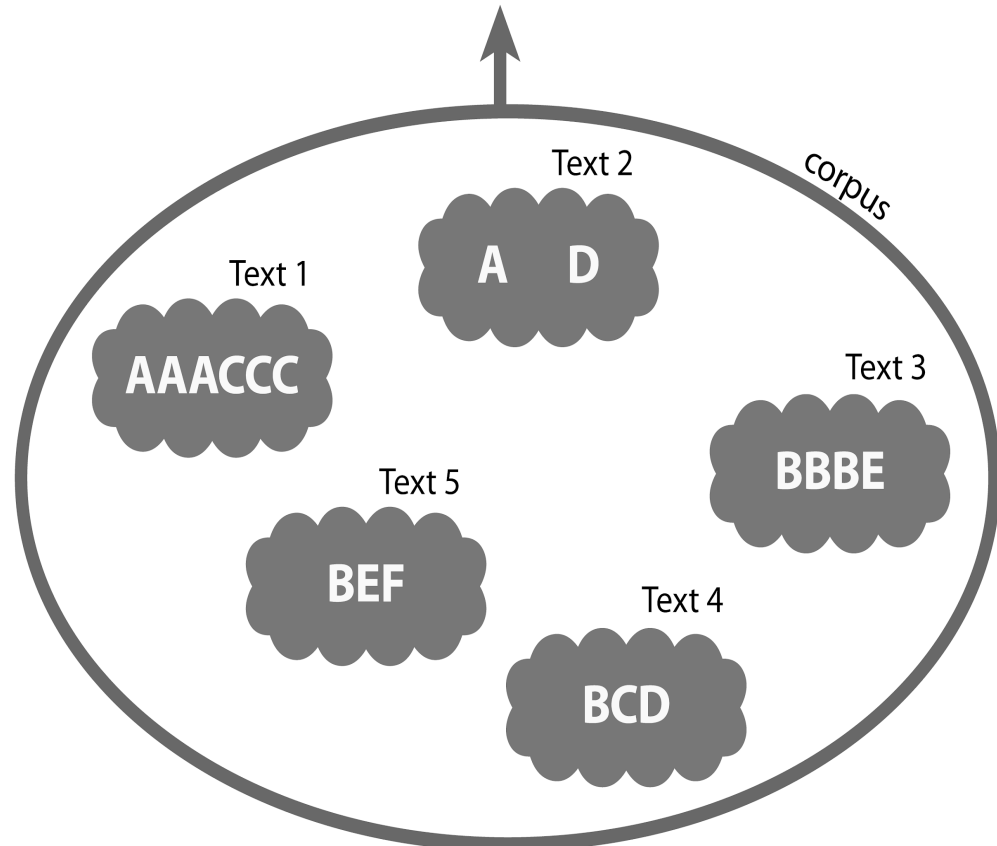
# Overview

- What Market Basket Analysis can do to facilitate large media data, i.e. corpus with lots of short texts (thanks to Josef Šlerka for pointing this method out to us!)
- Pilot study of Anti-system media class (in contrast to the center-right mainstream)
  - Data, the big picture showing 'flooding' in ANTS in contrast to mainstream
  - Narrative lines in anti-system media class uncovered via MBA
- Overarching ideas and our hypothesis for another article

# What to do after harvesting keywords (KWs) from the texts

Keyword analysis

KWs: **A** **B** **C**



The resulting keywords are without context

- Concordance analysis
- Collocations of KWs (or ngrams, clusters and other types of multi-word expressions)
- KW links
- Semantic grouping the KWs (grouping determined by researchers)

Representation of X (i.e. Ukraine, Russia)

# Market Basket Analysis produces associative links

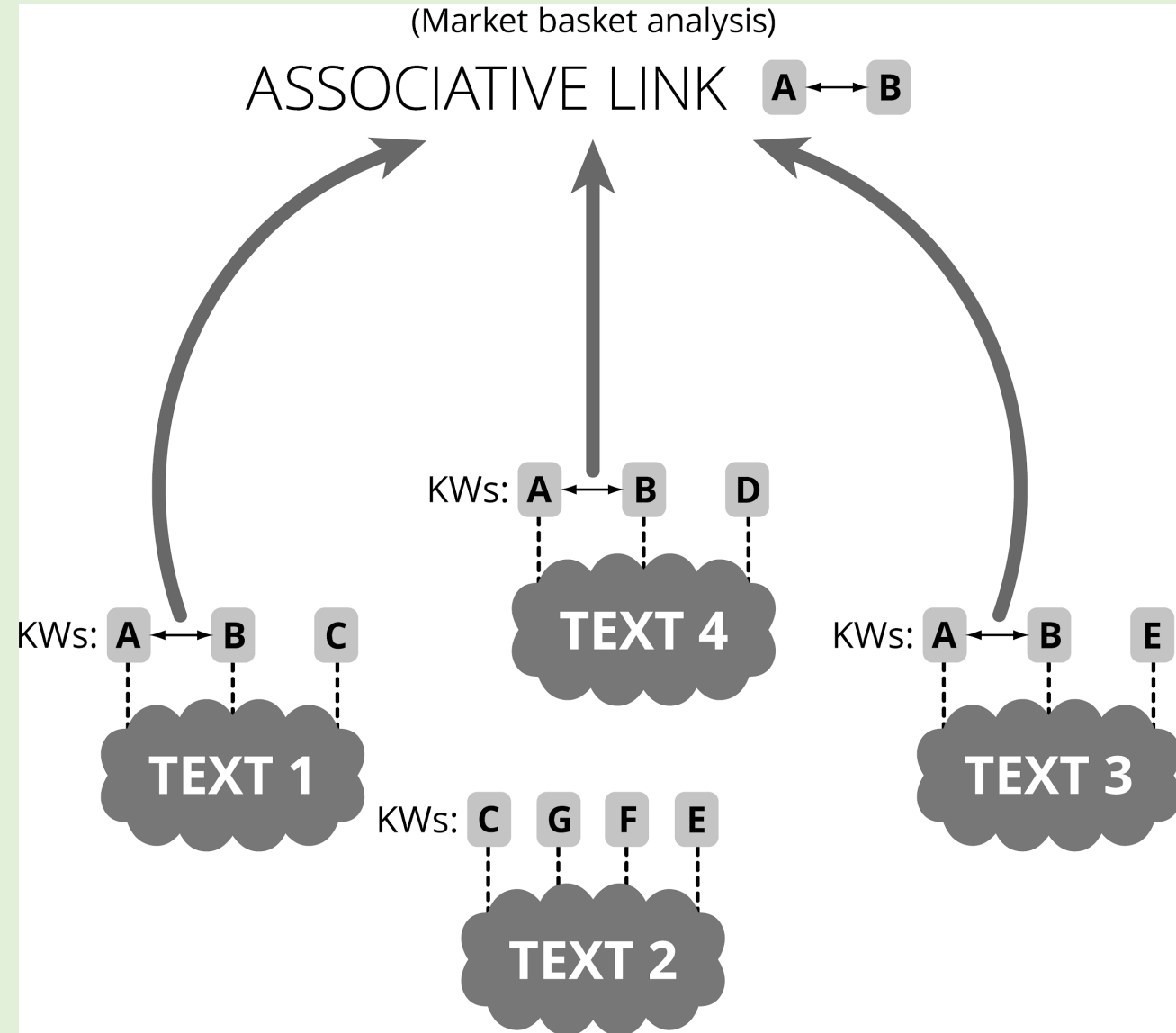
## 1. MBA

- informs us about some important aspects of who the customers are, points to their motivations and intentions.
- applied to discourse analysis: consistent ideas that permeate throughout the media class, overarching ideas and agendas regardless of seasonal topics
- the number of texts where KWs co-occur and the strength of associative links (ALs)

## 2. Discourse correlate of collocation:

- Collocation = which words cohabit within the vicinity

(cf. Cvrček and Fidler)



# Pilot

- 80 million-word WebMedia corpus (94 million tokens, incl. punctuation), Czech online web media portals (October 2017 to October 2018)
- Center-right mainstream (MS) vs. Anti-system (ANTS) media classes
- 19,700 KWs from ANTS, 24,300 KWs from MS
- MBA was applied to the results to yield 2-4-KW associative links: 60,400 ALs in MS and 284,800 ALs in ANTS
- Focus on ALs containing the KW *migrant*

# Highlights of the pilot

Media class	number of texts with <i>migrant</i> as a KW	number of ALs	median lift	median support
CR	256 (2.1 % of all CR texts)	235	14.5	0.0038
ANTS	472 (6.4 % of all ANTS texts)	1448	7.01	0.0039

- Outliers spread more widely in ANTS than MS
- "flooding" with topics
- Comparison between outlier ALs in ANTS and MS containing KWs unique to the media classes point to several narrative lines

# Narrative lines and samples 1

## **1. negative images of migrants**

*AL Africa, migrant, country → African*

“Afroislamic violent migrants and economic parasites”

Czech government’s collusion with the EU

## **2. current Czech government as lying to its public**

*AL Dublin, migration, migrant → IV*

“[Prime minister] Babiš's government in Luxembourg again did not oppose Dublin IV at the EU Council of Ministers, [...] ANO 2011 [the leading political party] says something else at home, but it keeps close pace with Berlin in voting in the EU! [...] Something stinks!”

# Narrative lines and samples 2

## **3. EU as an undemocratic organization.**

*AL migrant, minister → interior*

“Foreigners will flow into the Czech Republic regardless of what is said and promised to voters in the media. This can be stopped only by exiting the EU and NATO.”

## **4. The mainstream media hiding the truth (in contrast to ANTS)**

*AL migrant, applicant, country → asylum*

the Czech media is silent while it should signal a “high alert” (like ANTS) when the EU parliament started discussing “unilateral” redistribution of migrants



# Narrative lines and samples 3

## 5. Global conspiracy by transnational organizations

*AL EU, migrant, NGO → migration*

“[...] Arabic billboards started appearing in Prague! This is a blatant provocation of the nonprofit organization People in Need! The American neocons and their NATO are behind the migration to the EU!” advocating not only Czech EU-exit, but also NATO-exit, since the massive migration was said to have started and is being directed by Americans from 2011 with the help of NATO.

# Overarching ideas

Mainstream media is lying

Migrants are bad

The leading Czech politicians are lying

The EU and transnational organizations are damaging the member states



1. Readers should trust ANTS more than the mainstream media
2. Need to vote for more 'reasonable' politicians (like in Hungary)
3. Need to exit from the EU and NATO

Tentative hypothesis: these ideas (1-3 and their variations) might cohabit with other seasonal topics

# Text cited more in our article.

- Ex 1 Připomeňme si nyní, jak již vypadá již napůl africká Evropa dnes, a modleme se, aby za pár let nebyla afrikanizovaná zcela (<http://www.rukojmi.cz/clanky/zahranicni-politika/5051-diktat-volneho-trhu-je-v-africe-nutnou-podminkou-pro-zavedeni-zapadni-democraticke-diktatury-v-duchu-hesla-valka-je-dobry-byznys-nutne-je-vse-zlikvidovane-znovu-vybudovat-za-pomoci-darcu-a-veritelu-pod-zastitou-mmf-a-svetove-banky>)
- Ex 2 Babišova vláda v Lucemburku se opět na Radě ministrů EU nepostavila proti Dublinu IV, proti bylo jenom 7 zemí spolu s Itálií a Slovenskem, Česko mezi nimi nebylo. Hnutí ANO něco jiného říká doma, ale při hlasování v EU drží těsný krok s Berlínem! [...] Něco tady smrdí! (<https://aeronet.cz/news/babisova-vlada-v-lucemburku-se-opet-na-rade-ministru-eu-nepostavila-proti-dublinu-iv-proti-bylo-jenom-7-zemi-spolu-s-italii-a-slovenskem-cesko-mezi-nimi-nebylo-hnuti-ano-neco-jineho-rika-doma/>)
- Ex 3 Cizinci budou do ČR proudit bez ohledu na to, co se říká a slibuje voličům v médiích. Zastavit to jde pouze vystoupením z EU a z NATO, [...] A tak se hraje s voliči systém “na blba”, kdy v médiích padají ostrá slova o tom, že vláda nechce migraci, ale v téže chvíli nový ministr vnitra za ČSSD rozesílá obcím manuály, aby umožnily zápis cizinců do volebních seznamů a na kandidátky do podzimních voleb. (<https://aeronet.cz/news/cizinci-z-eu-bez-ceskeho-obcanstvi-budou-moci-na-podzim-nejen-volit-ale-dokonce-i-kandidovat-hamackovo-vnitro-rozeslalo-obecnim-uradum-metodicke-pokyny-ve-kterych-doporucuje-volebnim-komisim/>)
- Ex 4 Nejvyšší poplach! Evropský parlament dostal na stůl reformu Dublinského protokolu a česká média mlčí! (<http://zvedavec.org/komentare2018/01/7470-nejvyssi-poplach-evropsky-parlament-dostal-na-stul-reformup-dublinskeho-protokolu-a-ceska-media-mlci.htm>)
- Ex 5 [...] v Praze se začaly objevovat arabské billboardy! Je to nehorázná provokace neziskovky Člověk v tísni! Za migrací do EU stojí američtí neoconi a jejich NATO! [...] Vystoupení z EU a nevystoupení z NATO je cesta do pekel. [...] Migrační vlnu spustili a řídí Američané od roku 2011 za pomoci NATO. [...] Migrace a odpor lidí vůči Islámu dožene ve středoevropských zemích voliče k nenávisti vůči EU a k přesunu pod americké teze řízení a moci, pod křídla NATO jako americké úderné pěsti proti Rusku. Islámská migrace je tak mohutným plánem amerických neoconů na rozbití EU a pro zabránění globalistického spojení mezi Berlínem a Moskvou. (<https://aeronet.cz/news/foto-skoncily-volby-a-v-praze-se-zacaly-objevovat-arabske-billboardy-je-to-nehorazna-provokace-neziskovky-clovek-v-tisni-za-migraci-do-eu-stoji-americti-neoconi-a-jejich-nato/>)