

PHRASALEX II: Phraseological Approaches to Learner's Lexicography
22-23 July 2021

Discourse-oriented corpus studies as critical lexicography

Costas Gabrielatos

Edge Hill University

gabrielc@edgehill.ac.uk

<https://ehu.ac.uk/gabrielatos>

Critical Lexicography

- ❑ Focus: “The meanings of words which denote, represent, or reflect **politicized concepts and phenomena** – ideologically loaded items, totemic and socioculturally significant.”
- ❑ “Such words have been the **frequent focus** of linguistic investigations more widely, for example in **corpus-led studies from a discourse analytic perspective**, or sociological and cultural studies.”
- ❑ “**Ideologically positioned meaning is central to the concerns of critical lexicography.**”

(Moon, 2014: 85)

DOCS

- Core tenet: The critical examination of lexicogrammatical and discoursal patterns in context can help uncover attitudes towards particular entities (e.g. individuals, groups, issues).

Approach 1: Lexicography informing DOCS

- Dictionary definitions inform DOCS focus and/or are contrasted with DOCS findings.

Approach 2: DOCS as lexicography

- DOCS techniques and results seen as lexicographic work (e.g. Gabrielatos, 2011; Moon, 2014) → *emerging definitions*
- Can incorporate Approach 1: expected vs. unexpected (e.g. Krishnamurthy, 1996).

Parallel Lives?

- Scarce acknowledgement of the overlap in the objectives and methodology of DOCS and (critical) lexicography (cf. Gabrielatos, 2011; Moon, 2014).
- The approach to lexicography that Chen (2019) **proposes** is what DOCS have been **practising** since the early 1990s (see Gabrielatos 2020).

Collocation in DOCS

- ❑ Collocations “are often unavailable to intuition or conscious awareness. They can convey messages implicitly and even be at odds with an overt statement” (Hunston, 2002: 109).
- ❑ The frequent co-occurrence of two words (*node – collocate*) (Sinclair, 1991).
- ❑ Node and collocate need not be adjacent.
- ❑ Collocation need not be (part of) a (semi-) fixed expression.
- ❑ Collocations may be calculated within, or regardless of, sentence boundaries.
- ❑ cf. practice in lexicography – due to different objectives.
- ❑ Identifying collocations not an end in itself, but a way in.
- ❑ Collocational network (Phillips, 1989): Different nodes may share the same collocates.
- ❑ Intercollocation: Nodes may collocate themselves.

Three cases studies

Discourse presentation
(emerging definitions)
in newspapers

Asylum and Immigration

Islam

Autism

Discourses of refugees and asylum seekers in the UK Press, 1996-2006

CL team: Paul Baker, Costas Gabrielatos, Tony McEnery.

CDA team: Majid KhosraviNik, Michal Krzyzanowski, Ruth Wodak.

Articles: 175,139

Words: 139,510,037

Sub-corpora:

- per newspaper
- broadsheets/tabloids
- per year

Business
Evening Standard
Express
Daily Mail + Mail on Sunday
Daily Mirror + Sunday Mirror
Guardian + Observer
Herald
Independent + Independent on Sunday
Liverpool Echo
People
Star
Sun
Telegraph + Sunday Telegraph
Times + Sunday Times

	LDCE on CD-ROM (2003)	Oxford English Dictionary Online (2007)	Refugee Council
<i>refugee</i>	<p>Someone who has been forced to leave their country, especially during a war, or for political or religious reasons.</p> <p>Dictionary definitions of <i>refugee</i> make no mention of destination</p>	<p>One who, owing to religious persecution or political troubles, seeks refuge in a foreign country.</p> <p>A runaway; a fugitive from justice, etc. <i>rare</i>.</p> <p>Someone driven from his home by war or the fear of attack or persecution; a displaced person.</p>	<p>Someone whose asylum application has been successful and who is allowed to stay in another country having proved they would face persecution back home.</p> <p>Only the RC definition of <i>refugee</i> mentions the success of the asylum application</p>
<i>asylum seeker</i>	<p>Someone who leaves their own country because they are in danger, especially for political reasons, and who asks the government of another country to allow them to live there.</p>	<p>A person seeking refuge, esp. political asylum, in a nation other than his or her own.</p> <p>Both dictionary and RC definitions of <i>asylum seeker</i> make mention of asylum application</p>	<p>Someone who has fled persecution in their homeland, has arrived in another country, made themselves known to the authorities and exercised the legal right to apply for asylum.</p>

	LDCE on CD-ROM (2003)	Oxford English Dictionary Online (2007)	Refugee Council
<i>immigrant</i>	Someone who enters another country to live there permanently .	<ul style="list-style-type: none"> One who or that which immigrates; a person who migrates into a country as a settler. 	---
<i>migrant</i>	Someone who goes to live in another area or country, especially in order to find work .	<ul style="list-style-type: none"> A person who moves temporarily or seasonally from place to place A person who moves permanently to live in a new country, town, etc., esp. to look for work, or to take up a post, etc.; 	[<i>economic migrant</i>] Someone who has moved to another country to work .

- ❑ **Forced migration:** refugees and asylum seekers
 - ❑ **Voluntary migration:** immigrants and (economic) migrants
- (International Association for the Study of Forced Migration)

Dictionaries
asylum seeker:
a *refugee* who has applied for asylum

Temporal sequence:
refugee → *asylum seeker*

A refugee **need not become** an asylum seeker.

<i>refugee</i> :	potentially temporary
<i>asylum seeker</i> :	potentially permanent

Refugee Council
refugee:
a successful *asylum seeker*

Temporal sequence:
asylum seeker → *refugee*

A refugee **has been** an asylum seeker.

<i>refugee</i> :	potentially permanent
<i>asylum seeker</i> :	potentially temporary

If newspapers operate according to the dictionary definitions, then ...

- *asylum seekers*, seen as seeking permanent or long-term residence, would perhaps be expected to share a large proportion of collocates with *immigrants/migrants*.
- *refugees*, seen as probably transient, would be expected to show little overlap in collocates with *immigrants/migrants*.

Shared Consistent Collocates

↓	<i>refugees</i>	<i>asylum seekers</i>	<i>immigrants</i>	<i>migrants</i>
<i>refugees</i>		34%	33%	18%
<i>asylum seekers</i>	47%		46%	26%
<i>immigrants</i>	34%	40%		39%
<i>migrants</i>	38%	38%	79%	

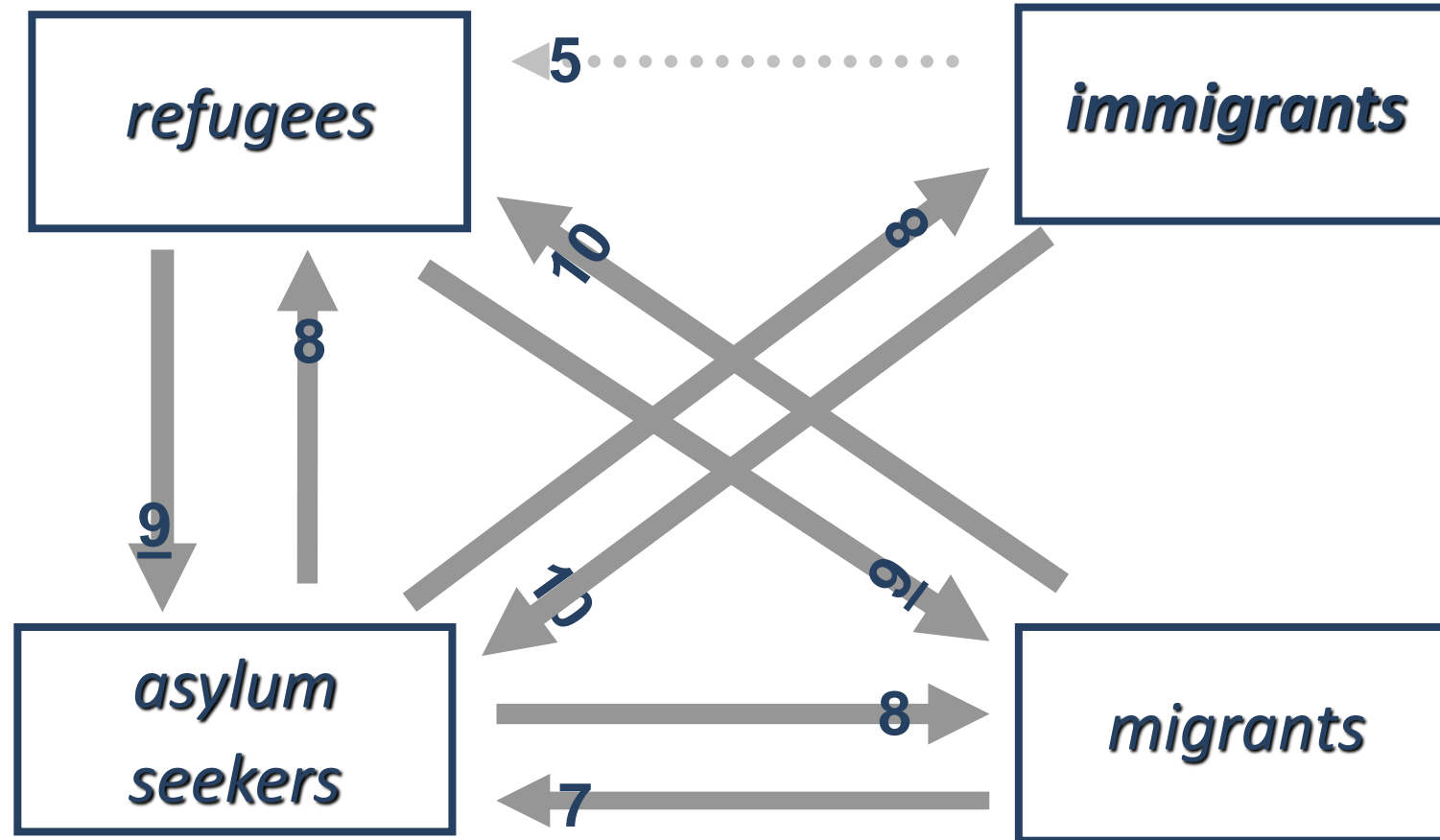
Similar proportions of shared collocates indicate similar discourses on two different types of migration (forced/voluntary).

Category	Definition and examples
Provenance / transit / destination	Used to refer to all forms of <i>RASIM</i> . Words referring to the country, region or continent that <i>RASIM</i> come from (provenance) e.g., Iran, Lebanon, Pakistan, Turkey, China; or the country, region or continent they want to go to (destination) e.g., UK, or the place that they temporarily stay, or are held, while waiting to be allowed go to their place of destination, or while waiting for their case to be decided (transit) e.g., France. As the same word may refer, in different texts, to more than one sub-category (e.g., <i>France</i> is both a country of destination and transit), these collocates were grouped together.
Number	Used mainly to refer to refugees, these are words denoting their large number (e.g., <i>flooding, pouring, streaming</i>).
Entry	Used mainly to refer to refugees, asylum seekers and migrants. Verbs denoting entry to the country of destination/transit (e.g., <i>arrive, come, enter</i>), verbs denoting their journey (e.g., <i>cross</i>), verbs denoting their (illegal) attempts to enter (e.g., <i>trying</i>), verbs denoting their being in transit (e.g., <i>waiting</i>), or the place of entry (e.g., <i>borders, Dover</i>).
Economic problem	Mainly used on asylum seekers, immigrants and migrants, these are words denoting either that they are a financial burden , because they receive state benefits (e.g., <i>benefits, claiming, receive</i>), or a financial threat , because they compete for jobs with existing citizens (e.g., <i>jobs, working</i>).
Residence	Mainly used on refugees and asylum seekers. Words denoting their residence during transit (e.g., <i>camp, shelter, temporary</i>), or in the destination country (e.g., <i>housed, settled</i>), or verbs used to express opposition to the latter (e.g., <i>allowed, granted</i>).
Return / repatriation	Mainly used on refugees and asylum seekers. Words referring to <i>RASIM</i> returning , or being made to return , to their country (e.g., <i>back, refused, return, sent</i>)
Legality	Mainly used on asylum seekers and immigrants, these are words concerning the legal status of <i>RASIM</i> . They are either direct, that is, attributive (e.g., <i>bogus, genuine, illegal</i>), or indirect, that is, they imply their illegality (e.g., <i>caught, detained, smuggled</i>).
Plight	Mainly used to refer to refugees. Words referring to the situations that necessitated their leaving their country (e.g., <i>fear, forced, persecution</i>), the manner of their leaving (e.g., <i>escape, fleeing</i>), their current/recent state (e.g., <i>displaced, homeless</i>), or their current needs (e.g., <i>aid</i>).

Shared c-collocates: Categories

↓	<i>refugees</i>	<i>asylum seekers</i>	<i>immigrants</i>	<i>migrants</i>
<i>refugees</i>		ENTRY NUMBER ECON. BURDEN RETURN	ENTRY RESIDENCE	ENTRY
<i>asylum seekers</i>	ENTRY PLIGHT NUMBER RETURN		ENTRY LEGALITY PTD RESIDENCE	ENTRY
<i>immigrants</i>	ENTRY RESIDENCE PLIGHT NUMBER	ENTRY PLIGHT RESIDENCE LEGALITY		ENTRY ECON. THREAT
<i>migrants</i>	ENTRY RESIDENCE PLIGHT PTD	PLIGHT	PTD ENTRY RESIDENCE ECON. THREAT LEGALITY	

Consistent intercollocations of *RASIM*



Figures indicate the number of annual sub-corpora in which terms intercollocate.

Racism discourse sandwich

WHAT a horrific, callous man Perry Wacker is to let those 58 Chinese **migrants** suffocate in the rear of his truck. If our Government had stood firm and made it difficult to enter Britain - turning migrants back instead of looking after them - they would not try to smuggle themselves here. Then this tragic waste of life and the anguish of the people who found them might not have happened. The manslaughter charge should have been shared by the Government for not sorting out the problem.

[Letter, *The Sun*, June 2001]

The representation of Islam and Muslims in the UK press, 1998-2009

Paul Baker, Costas Gabrielatos, Tony McEnery

Articles: *200,000*

Words: *143 million*

Spelling normalisation

Sub-corpora:

- per newspaper
- per year (1998-2009)
- broadsheets/tabloids
- political orientation

Business

Daily Express + Sunday Express

Daily Mail + Mail on Sunday

Daily Mirror + Sunday Mirror

Guardian + Observer

Independent + Independent on Sunday

People

Daily Star + Daily Star Sunday

Sun

Telegraph + Sunday Telegraph

Times + Sunday Times

Corpus tools and methodology

Tool

- Sketch Engine (Kilgarriff et al., 2004)
 - Originally developed for lexicographical purposes

Methodology

- Detailed wordlist analysis, keyword analysis, word sketches
- Co-textual analysis

Word Sketch

- Collocates of a word within a grammatical construction
 - *Muslim_Adj* + Noun collocates

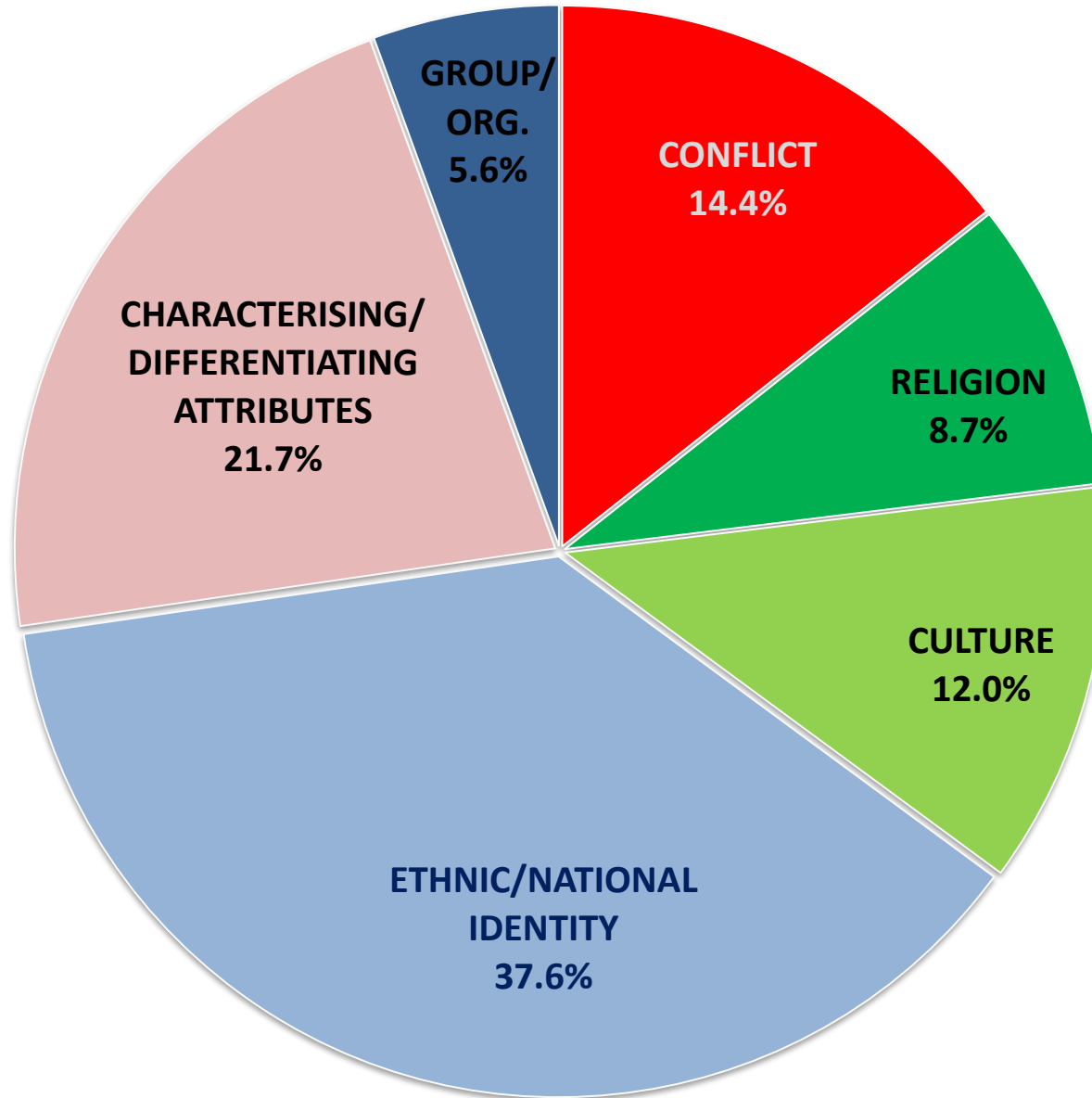
Muslim Islam-UK freq = 89834

Sketch of *Muslim* used as an adjective

and/or	24853 3.0	ADJ+n	84671 6.9	adv+ADJ	2648 2.4	v+ADJ	1858 1.1	N+at-i	6 0.0
British	1660 8.32	community	7676 10.4	predominantly	583 11.99	be	1709 3.82	-a-year	1 6.65
Arab	1113 9.27	world	4928 9.19	mainly	467 10.93	become	23 2.76	beginning	1 1.68
young	1035 8.39	woman	4484 9.07	largely	238 9.14	look	14 2.61	%	1 0.46
Jewish	879 9.11	country	3864 8.74	mostly	153 9.25	bear	12 3.65	N+from-i	2 0.0
Christian	802 9.06	leader	3217 8.8	not	143 3.03	remain	7 2.15	birth	1 1.62
radical	722 9.21	cleric	2279 9.36	overwhelmingly	115 9.86	appear	5 2.26	age	1 0.13
holy	692 9.19	man	2158 7.5	only	80 4.83	turn	5 1.64	N+to-i	2 0.0
first	667 7.18	group	1837 7.74	strictly	40 8.07	win	4 1.51	agnostic	1 6.04
other	645 6.86	population	1799 8.89	even	40 4.39	call	4 0.84	N+in-i	38 0.0
□□□	517 6.05	extremist	1710 8.81	devoutly	38 8.77	compare	3 3.16	origin	3 3.22
many	515 7.06	school	1220 7.67	exclusively	30 7.85	feel	3 1.01	name	3 0.57
traditional	445 8.35	state	997 7.35	especially	30 7.23	bring	3 0.7	field	2 1.73
large	444 7.78	girl	966 7.8	about	28 4.21	lead	3 0.45	court	2 0.15
moderate	376 8.61	family	901 7.05	also	28 2.22	include	3 0.2	makeup	1 5.32
Bosnian	323 8.3	faith	823 7.68	now	24 2.63	mind	2 3.84	squadron	1 5.03
strict	312 8.37	nation	805 7.66	back	22 3.59	mix	2 3.48	census	1 4.47
local	312 6.99	organisation	696 7.45	entirely	20 5.93	recognise	2 2.36	ethic	1 3.62
fundamentalist	277 8.13	youth	630 7.64	n't	20 0.74	stay	2 1.21	th-century	1 2.22
senior	254 7.26	fanatic	612 7.71	just	19 2.45	think	2 0.34	neighbourhood	1 2.16
religious	232 6.51	student	579 7.18	particularly	18 5.5	want	2 0.32	bid	1 1.6
populous	229 8.18	soldier	519 6.53	heavily	17 5.72	mint	1 4.06	design	1 1.55
militant	224 7.56	child	505 6.2	staunchly	16 7.47	disenfranchise	1 3.96	prayer	1 0.9
female	221 7.57	people	489 4.91	nominally	16 7.46	fast	1 3.64	character	1 0.43
devout	212 7.97	terrorist	467 6.71	strongly	16 6.07	modernise	1 3.55	N+for-i	5 0.0
black	205 6.7	pupil	452 7.2	specifically	14 6.63	articulate	1 3.37	chunk	1 3.65
Asian	201 7.16	parent	441 6.87	all	13 3.0	prohibit	1 3.34		
Sikh	185 7.78	prisoner	440 6.95	as	13 2.55	dare	1 3.33		
mainstream	171 7.6	scholar	425 7.21	very	12 2.08	exile	1 3.19		
-year-old	169 6.78	fundamentalist	424 7.21	so	12 1.89	sack	1 2.96		
only	139 6.05	area	413 6.31	too	11 2.76	slaughter	1 2.92		

Collocates: categories and sub-categories		Examples of noun collocates
CONFLICT		<i>extremist, fanatic, terrorist, fundamentalist</i>
RELIGION		<i>cleric, faith, festival, preacher</i>
CULTURE	Social practices	<i>dress, culture, name, tradition</i>
	Education	<i>school, teaching, education, college</i>
	View/attitude/emotion	<i>opinion, anger, voice, attitude, grievance</i>
ETHNIC/ NATIONAL ENTITY	Population	<i>community, population, nation, world</i>
	Area/country	<i>country, state, area, region, land</i>
	Governance	<i>leader, voter, MP, government, ruler</i>
CHARACTERISING/ DIFFERENTIATING ATTRIBUTES	Age/sex	<i>woman, man, girl, youth, child, teenager</i>
	Family/relationship	<i>family, parent, brother, friend, wife</i>
	Occupation/role	<i>officer, patient, doctor, worker, assistant</i>
	Ethnicity/race/nationality	<i>Briton, Albanian, Malay, Arab</i>
	Other	<i>house, shop</i>
GROUP / ORGANISATION		<i>group, organisation, association, charity</i>

Muslim-Adj: Noun collocates - tokens

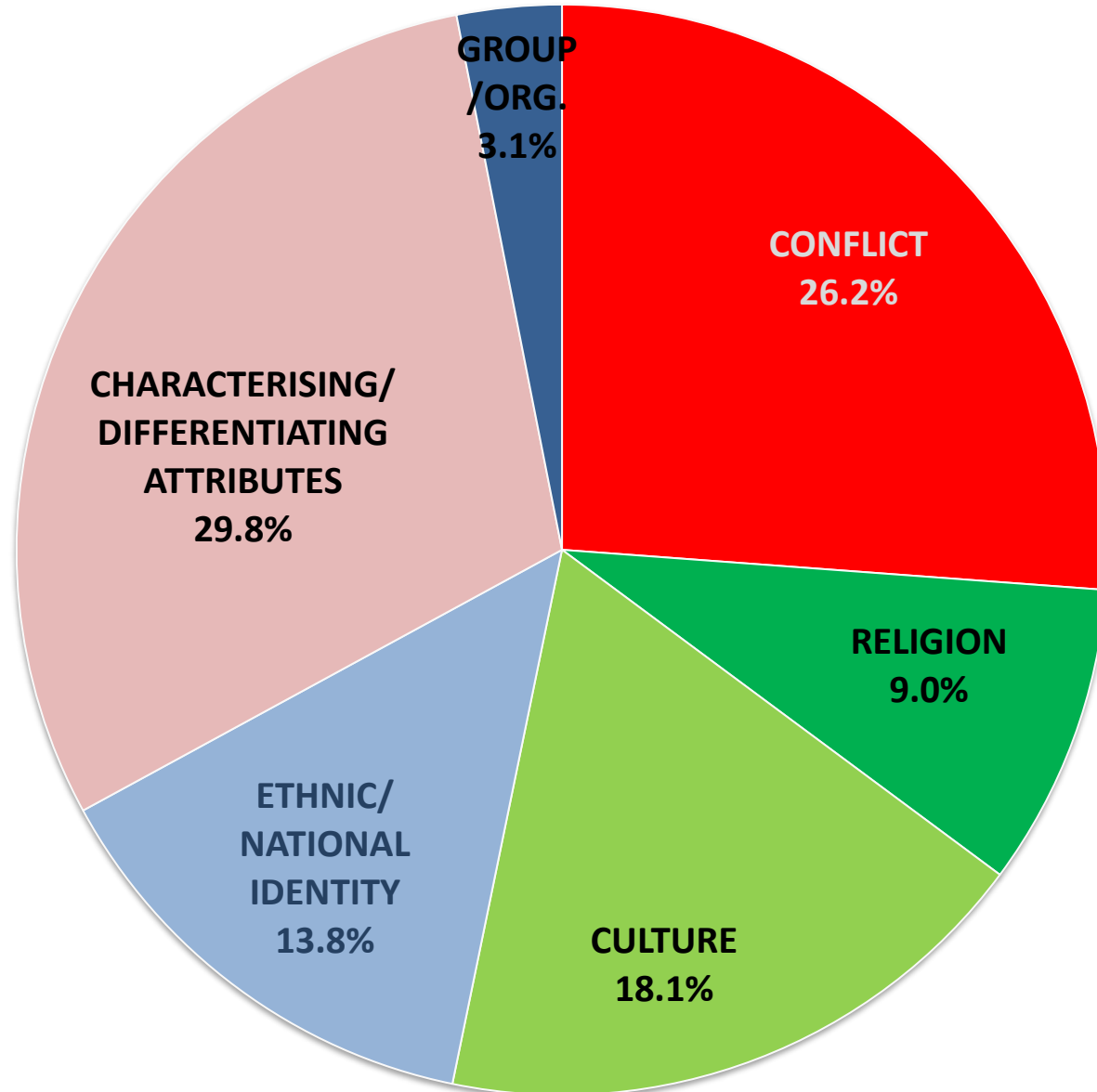


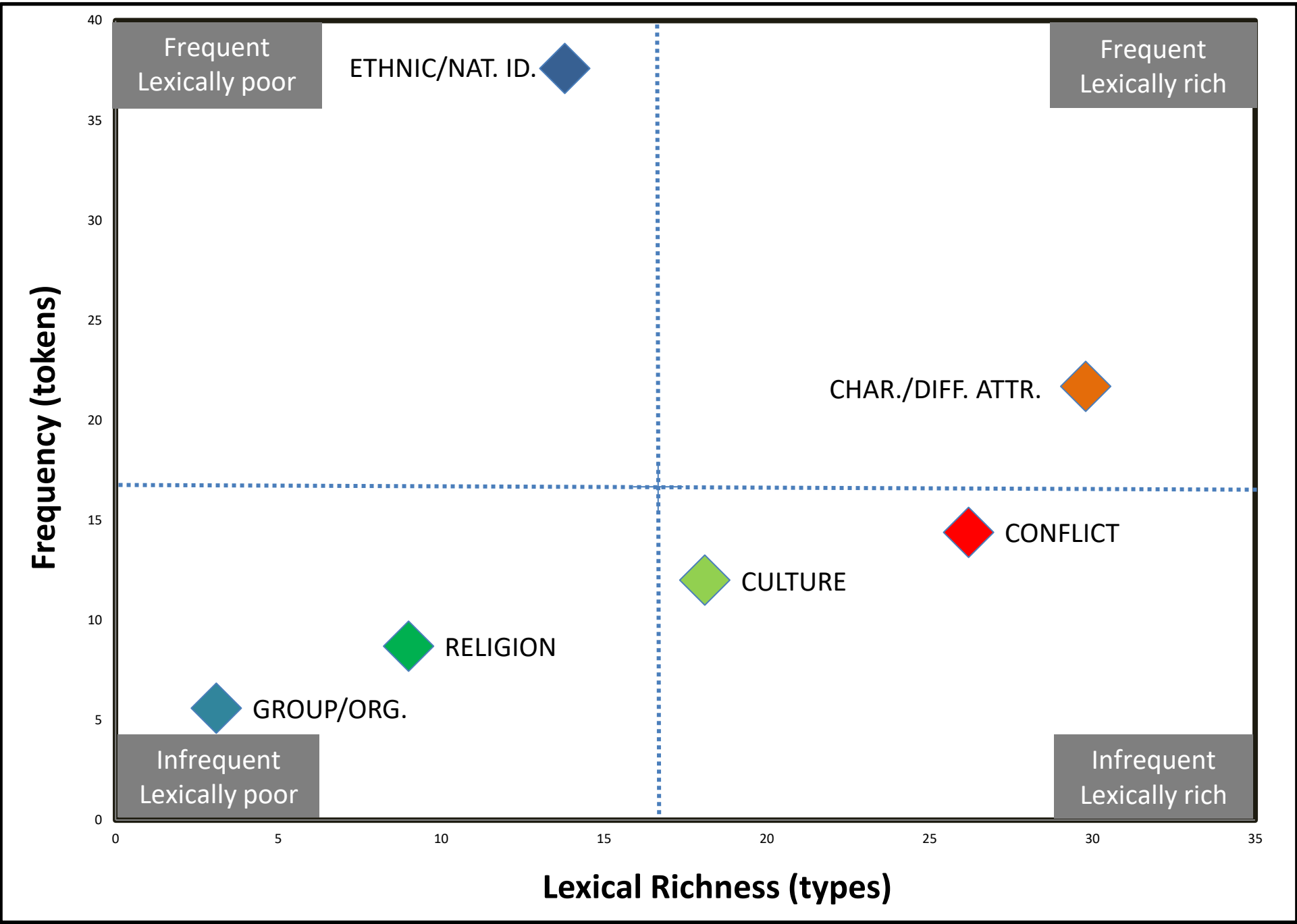
*The distribution in terms of tokens may
be due to some very frequent types*



Distribution in terms of types

Muslim_Adj: Noun collocates - types





Emerging patterns

- Muslims are usually presented as a homogeneous group.
- *Muslim_Adj* is used much more often as a national/ ethnic/cultural attribute than a religious one → they are 'different'.
- The direct mention of armed/social conflict is both very frequent and lexically rich.
- Other uses of *Muslim_Adj* are usually embedded within discussions of conflict.

(Baker, Gabrielatos, McEnery, 2013a, 2013b)

Extended co-text of collocations helps uncover implicit definitions

A LEADING barrister says Sharia law in Britain is "inevitable", despite 95% of Daily Star readers insisting it is wrong in our phone poll yesterday. Most of our readers believe that **Britons should not have to live under the controversial Muslim code** in their own country, which includes stoning to death for adultery and cutting off hands for theft. But former Chairman of the Bar Council Stephen Hockman said there was no reason why Sharia law could not be applied here. He said : "Given the world situation and our own substantial Muslim population, **it is vital that we now look at ways to integrate Muslim culture into our own traditions.**

[Daily Star, 05.07.2008]

Portrayals of Autism in the British Press 2011-2020

Costas Gabrielatos, Themis Karaminis, Ursula Maden-Weinberger

Articles: 24,000

Words: 19 million

Sub-corpora:

- per newspaper
- per year (2011-2020)
- broadsheets/tabloids
- political orientation

Daily Express + Sunday Express

Daily Mail + Mail on Sunday

Daily Mirror + Sunday Mirror

Guardian + Observer

Independent + Independent on Sunday

Daily Star + Daily Star Sunday

Sun

Telegraph + Sunday Telegraph

Times + Sunday Times

Autism: sources of definitions

- Five general online dictionaries
 - Cambridge, Collins, Longman, Macmillan, Merriam-Webster, Oxford
- Autism Diagnosis Criteria: DSM-5 ('deficit' approach)
(<https://www.autismspeaks.org/autism-diagnosis-criteria-dsm-5>)
- National Autistic Society (<https://www.autism.org.uk>)
- Autistica (<https://www.autistica.org.uk>)
- 'diversity' approach

Corpus analysis: Emerging definitions (1)

What is autism?

-- Both 'deficit' and 'diversity' approaches reflected:

- *developmental, common, complex, lifelong, genetic, challenging, neurological*
- *disability, disorder, condition*
- [manifests in] *difficulties, problems, special needs, issues, challenges*
- [affects] *social interaction, behaviour, communication, sensory experiences*

(Maden-Weinberger et al., 2021)

Corpus analysis: Emerging patterns (2)

Who has autism?

- *son, boy, child, adults*
- Only from 2015 onwards is *girl* a collocate of *autism*
- Since 2018: *daughter, babies, students, pupils, teenagers*

Autism is mentioned together with:

- *learning disabilities, ADHD*
- *mental health disorders, schizophrenia, depression, anxiety, bipolar disorder, psychiatric conditions, dementia, Tourette's syndrome*
- *epilepsy, cerebral palsy, diabetes, asthma, multiple sclerosis*

(Maden-Weinberger et al., 2021)

Corpus analysis: Emerging patterns (3)

No/little agency:

- Autistic people (regardless of age) are not presented as representing themselves, but they are spoken about.

Mothers are the carers

- 5 times more frequent collocates than fathers

Autistic children are presented as

- "attributes" of mother
- represented by mother
- the reason for mother's actions/activism/advocacy
- a 'burden'/'problem' for carers/society

(Maden-Weinberger et al., 2021)

A burden for carers

- Having spent much of her childhood as a carer to her disabled father, the discovery of her son's **autism** was, Lucy says, a **heavy blow**. (*Daily Mail*, July 2011)

Discourse Comorbidities + Nature of autism

- More **mental illness** 'in children of older dads': Greater risk of **autism**, ADHD and schizophrenia. (*Guardian*, February 2014)

No agency - parent is representative

- He also established the Strategic Autism Initiative, which he ran with Polly Tommey, a British mother with an **autistic** son. (*Independent*, May 2018)

'Cure'

(Expressed directly or indirectly – sometimes because of choice of words, e.g. 'therapies', or the polysemy of 'treatment')

- Thanks to special classes and **therapies** and indeed her father's tireless work raising **autism** awareness, Mia learned to speak when she was seven years old and eventually she made the transition into regular school. (*Daily Mail*, May 2018)
- We are meeting in a west London cafe to talk about her new project, a centre for the **treatment** of children with autism. (*Times*, August 2018)

References

- Baker, P., Gabrielatos, C. & McEnery A. (2013) *Discourse Analysis and Media Attitudes: The representation of Islam in the British Press*. Cambridge University Press.
- Baker, P., Gabrielatos, C. & McEnery, T. (2013) Sketching Muslims: A corpus-driven analysis of representation around the word “Muslim” in the British press, 1998-2009. *Applied Linguistics*, 34(3), 255–278.
- Baker, P., Gabrielatos C., KhosraviNik, M., Krzyzanowski, M., McEnery, T. & Wodak, R. (2008) A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273-305
- Chen, W. (2019) Towards a discourse approach to critical lexicography. *International Journal of Lexicography*, 32(3), 362–388
- Gabrielatos, C. (2020, August 29) Bibliography of Discourse-Oriented Corpus Studies. Retrieved from <http://ehu.ac.uk/docsbiblio>.
- Gabrielatos, C. (2011). Collocational approaches to critical discourse studies: A case of critical corpus lexicography. *International Symposium On The Sociology Of Words: Lexical Meaning, Combinatorial Potential and Computational Implementation*. LACELL, University of Murcia, Spain, 1-2 December 2011.
- Gabrielatos (2007) Selecting query terms to build a specialised corpus from a restricted-access database. *ICAME Journal* 31, 5-43.
- Gabrielatos, C. & Baker, P. (2008) Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK Press 1996-2005. *Journal of English Linguistics*, 36(1), 5-38.
- Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge University Press.
- Kilgarriff, A., P. Rychly, P. Smrz & D. Tugwell. (2004) The Sketch Engine,. Proceedings of Euralex. 105–116.
- Krishnamurthy, R. (1996) Ethnic, Racial and Tribal: The language of racism? In Caldas-Coulthard, C.R. & Coulthard, M. (eds.) *Texts and Practices: Readings in Critical Discourse Analysis*. Routledge. 129-149.
- Maden-Weinberger, U., Gabrielatos, C., Beattie, G. & Karaminis, T. (2021) Implicit attitudes towards autism in the British Press. *Autistica Research Festival*, 12-16 July 2021.
- Moon, R. (2014) Meanings, ideologies, and learners’ dictionaries. In Abel, A., Vettori, C. & Ralli, N. (eds.) *Proceedings of the XVI EURALEX International Congress: The user in focus*. 15-19 July 2014. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism. 85-105. [[e-copy](#)]
- Phillips, Martin. (1989) *Lexical Structure of Text*. Discourse Analysis Monograph no. 12, English Language Research, University of Birmingham, UK.
- Sinclair, J.McH. (1991) *Corpus Concordance Collocation*. Oxford University Press.