

## Using collocation analysis to reveal the construction of minority groups: The case of refugees, asylum seekers and immigrants in the UK press

Paul Baker, Tony McEnery and Costas Gabrielatos

Department of Linguistics and English Language, Lancaster University  
Lancaster LA1 4YT, United Kingdom,  
j.p.baker@lancaster.ac.uk; a.mcenery@lancaster.ac.uk; c.gabrielatos@lancaster.ac.uk

Refugees, asylum seekers, and immigrants (henceforth RASIM) coming into the UK have attracted increased press attention (Greenslade, 2005). As their representation in the press can construct their identity (Duffy and Rowden, 2005: 6, in Greenslade, 2005: 7), the discourses surrounding these groups have been the focus of linguistic studies (e.g. ter Wal, 2002).

This paper reports on the ESRC funded project, 'Representation of refugees and asylum seekers in UK newspapers 1996-2005'. Although the project combines critical discourse analysis and corpus linguistics approaches, the paper aims to exemplify the contribution of corpus research to discourse analysis (cf. Koller & Mautner, 2004; Orpin, 2005; Sotillo & Wang-Gempp, 2004). The study used a corpus of 140 million words, comprising 175,000 articles from 15 UK newspapers, spanning 1996-2005 (see Gabrielatos, 2007). Specifically, the paper reports on the collocational analysis, which adopted the methodology in Baker & McEnery (2005) and McEnery (2006). An added methodological notion, akin to *key keywords* (Scott, 2004: 115), is that of *consistent collocates*, i.e. words which are collocates in at least seven out of the ten annual sub-corpora.

Collocates can contribute to "a semantic analysis of a word" (Sinclair, 1991: 115-116). Also, as "they can convey messages implicitly and even be at odds with an overt statement" (Hunston, 2002: 109), they are a suitable vehicle for the discursive presentation of a group (Baker, 2006). The analysis also makes use of the related notions of *semantic prosody* (Louw, 1993: 157), *semantic preference* (Stubbs, 2001: 65), and *discourse prosody* (ibid.: 65-66). The examination of collocation patterns has revealed systematic semantic associations, which map onto the CDA notions of *topos* (Reisigl & Wodak, 2001: 74-76) and *topic* (Sedlak, 2001: 129-130), as well as metaphors commonly employed in racist discourse (van der Valk (2000: 234). Arguably, these patterns reveal elements of the underlying discourses relating to RASIM.

Corpus Linguistics 2007, 28-30 July 2007

## Using collocation analysis to reveal the construction of minority groups

The case of refugees, asylum seekers  
and immigrants in the UK press

Paul Baker, Tony McEnery, Costas Gabrielatos  
Lancaster University

## The RASIM Corpus (1996-2005)

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*

## Informing concepts

- **Collocation**
  - collocates can provide “a semantic analysis of a word” (Sinclair, 1991: 115-116).
  - collocation patterns “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).
- **Collocational network** (Phillips, 1989)
  - Shared collocates and intercollocation of *RASIM*.
- **Semantic preference** (Stubbs, 2001: 65)
  - Indexes the concepts, topics and issues routinely associated with *RASIM*.
- **Semantic prosody** (Louw, 1993: 157)
- **Discourse prosody** (Stubbs, 2001: 66)
  - Index attitudes and evaluative stance towards *RASIM*.
- **Topoi**
  - “Conclusion rules that connect the argument with the conclusion” (Reisigl and Wodak, 2001: 74-76).
  - Used to refine categories emerging from concordance analysis.

## A problem

- The *RASIM* corpus is diachronic.
- Extracting collocates from the whole corpus can be expected to include a large number of ‘seasonal collocates’.
- Confirmed by collocational analysis of annual sub-corpora.
- Cumulative breakdown of *RASIM* collocates present in up to five annual sub-corpora:

Term	1 year	1-2 years	1-3 years	1-4 years	1-5 years
<i>refugees</i>	56.7%	73.3%	81.9%	87.0%	92.3%
<i>asylum seekers</i>	58.2%	76.6%	83.7%	88.9%	92.8%
<i>immigrants</i>	50.8%	68.8%	78.2%	83.5%	88.3%
<i>migrants</i>	62.3%	77.4%	86.1%	90.2%	94.7%

Most collocates are ‘seasonal’.

## Solution: Consistent Collocates

- Scott (1998: 97-98) uses the notion of *consistency* in relation to word lists and keywords.
- A *consistency analysis* shows the number of texts or sub-corpora that a word is found in.
- Collocates of *RASIM* were calculated for each annual sub-corpus.
- *Consistent collocates* were deemed those present in at least seven of the ten sub-corpora.
- They indicate core elements of meaning, semantic associations and semantic preferences/prosodies.

Term	Ratio of c-collocates
<i>refugees</i>	7.7%
<i>asylum seekers</i>	4.5%
<i>immigrants</i>	8.9%
<i>migrants</i>	4.0%

## Consistent Collocates: Calculation

- Collocates calculated in each annual sub-corpus (using *WordSmith Tools*, Scott, 1999).
- Annual collocates tabulated, and c-collocates established manually.
- Category inclusion established through concordance analysis.
- Span:  $\pm 5$  (within sentence boundaries).
- MI determines strength of collocation, but not statistical significance (McEnery, 2006: 22).
- LL determines statistical significance.
- $\Rightarrow$  Cut-off point for c-collocates:  $MI \geq 3$  and  $LL \geq 6.63$  ( $p \leq 10^{-1}$ ).
- Minimum frequency of collocates: 5 (3 for *migrants*).
- $\Rightarrow$  Minimum frequency of c-collocates: 35 (21 for *migrants*).

## Findings from relevant study

(Baker & McEney, 2005)

*Refugees* commonly described in terms of:

- provenance
- temporary residence
- destination
- quantity
- movement
- plight

*Asylum seekers* frequently linked to:

- issues of legislation related to immigration
- problems allegedly stemming from their presence in the host country
- illegal / unsuccessful / would-be immigrants

## Focus

- C-collocates of *refugees*, *asylum seekers*, *immigrants*, *migrants* (RASIM).
- Common c-collocates of RASIM.
- Categorisation of c-collocates informed by *topoi*, *topics* or *metaphors* recognised in CDA (Reisgl & Wodak, 2001; Sedlak, 2000; van Leuven, 1996; Wodak & van Dijk, 2000).
- Overlap of categories between RASIM.
- C-intercollocations between the terms.

## Definitions

	Longman Dictionary of Contemporary English (2003)	Refugee Council
<i>refugee</i>	Someone who has been forced to leave their country, especially during a war, or for political or religious reasons.	Someone whose asylum application has been successful and who is allowed to stay in another country having proved they would face persecution back home.
<i>asylum seeker</i>	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.	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.
<i>immigrant</i>	Someone who enters another country to live there permanently.	-----
<i>migrant</i>	Someone who goes to live in another area or country, especially in order to find work.	[ <i>economic migrant</i> ] Someone who has moved to another country to work.
	<i>refugee</i> → <i>asylum seeker</i>	<i>asylum seeker</i> → <i>refugee</i>
	A refugee need not be/become an asylum seeker.	A refugee has been an asylum seeker.

## Plight

<i>refugees</i>	<i>asylum seekers</i>	<i>immigrants</i>	<i>migrants</i>
aid desperate displaced driven escape fear fighting fled flee fleeing forced homeless homes leave persecution torn	fled  fleeing forced	fled [!]	fleeing [!]

## Residence

<i>refugees</i>	<i>asylum seekers</i>	<i>immigrants</i>	<i>migrants</i>
allowed asylum camp camps granted  seek seeking settle settled shelter sheltering stay  waiting	accommodation allowed  granted housed housing live living  stay temporary	living  seeking [!] settle settled  stay	allowed asylum [!]  seeking [!]  

## Economic burden/threat

<i>refugees</i>	<i>asylum seekers</i>	<i>immigrants</i>	<i>migrants</i>
allowed  economic	allowed benefits claiming economic entitled  receive	jobs  working	allowed  jobs  work
economic burden		economic threat	

## Shared c-collocates: proportion (non-directional)

Pairs of terms		% of shared c-collocates
<i>immigrants</i>	<i>migrants</i>	59.0%
<i>asylum seekers</i>	<i>immigrants</i>	43.0%
<i>refugees</i>	<i>asylum seekers</i>	40.5%
<i>refugees</i>	<i>immigrants</i>	33.5%
<i>asylum seekers</i>	<i>migrants</i>	32.0%
<i>refugees</i>	<i>migrants</i>	28.0%

## Shared c-collocates: proportion (directional)

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



### Shared c-collocates: categories

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

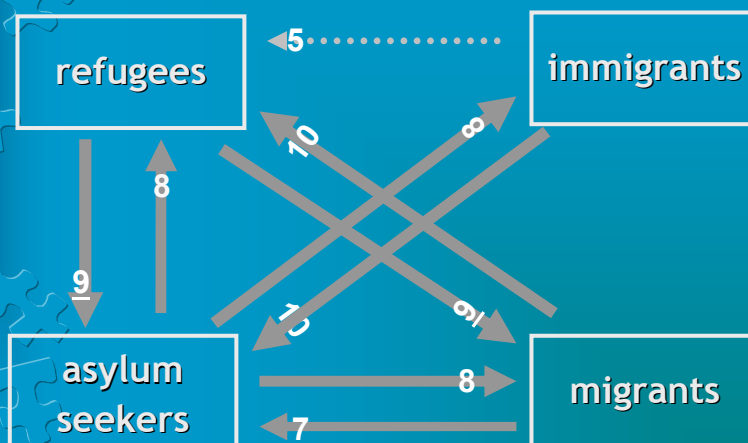
### C-collocates indexing different categories: The case of *allowed*

Entry	FRANCE finally closed a loophole yesterday which has <i>allowed</i> thousands of <i>illegal immigrants</i> to sneak into Britain on Eurostar trains. ( <i>Daily Star</i> , 5 February 2002).
Residence	Jack Straw faced embarrassment last night as it emerged that almost one-third of new <i>asylum seekers</i> have been <i>allowed</i> to remain in Britain. ( <i>Daily Mail</i> , 26 November 1998)
Economic burden / threat	<u>Burden</u> : But <i>illegal immigration</i> and <i>bogus asylum seekers</i> have been <i>allowed</i> to pour into Britain to live off our taxes. ( <i>The Sun</i> , 21 April 2005). <u>Threat</u> : Job firms in the South have been inundated with pleas for jobs from <i>illegal migrants</i> who are <i>allowed</i> to work. ( <i>Sunday Mirror</i> , 16 April 2000)
Return	Only a small number of <i>refugees</i> have been <i>allowed</i> to return home, and there has been muted progress in merging Serb-controlled areas with those of the Muslim-Croat federation. ( <i>The Times</i> , 25 September 1997)

## Shared c-collocates indexing interchangeable use of *RASIM*: The case of *trying*

- English, Welsh and Scottish Railways (EWS) warned last night that it cannot, and will not, endure another six months of lost revenue - (pounds) 10m so far - arising from security problems caused by asylum seekers in France. EWS services to and from Europe have fallen by 60% since November because so many refugees are *trying* to board trains bound for Britain via the Channel tunnel. (*The Herald*, 7 May 2002)
- CHANNEL Tunnel security came under scrutiny last night after 44 illegal immigrants were intercepted *trying* to reach Dover. The desperate asylum seekers walked seven miles in complete darkness before being caught. (*The Mirror*, 31 August 2001)
- All the passengers were illegal immigrants *trying* to make their way to Greece. Survivors identified them as Pakistanis, Moroccans and Bangladeshis. Mr Dokuzoglu said Indian and Afghan refugees were also believed to be on board. (*The Guardian*, 2 January 2001)

## Consistent intercollocations of *RASIM*



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

## Intercollocations : Contexts

### □ Problem/issue (here: economic problem):

A CRACKDOWN on London's black market jobs racket for asylum seekers and illegal immigrants will be launched today by David Blunkett as he unveils sweeping reforms of the immigration system. (*The Evening Standard*, 7 February 2002)

### □ Plight

- The scheme helps the homeless, refugees and asylum-seekers, young people leaving care, the elderly, and women fleeing domestic violence. (*Liverpool Daily Echo*, 26 August 2005)

### □ Scepticism about RAS being genuine

- Mr Blunkett must find a much faster way of sorting out genuine refugees from economic immigrants. (*The Express*, 5 September 2001)

### □ Confusion/conflation of use

- Gradually - very gradually - a policy solution seems to be emerging. Countries start by allowing entry to job-creating entrepreneurs. Then they admit workers with skills that are in short supply. They also accept, perhaps for a limited period, large numbers of unskilled workers needed to keep hospital floors swept and construction projects going. Then they allow in some immigrants who are genuine asylum seekers. (*The Times*, 30 July 2000)

## Conclusions

- The significant overlap of c-collocates of RASIM, indicates the use of these terms in the same/similar contexts, and with similar attitudes.
- Large proportion of shared c-collocates and strong intercollocations can be interpreted as an indication of overlap in their usage, and, in some cases, their use as near synonyms.
- The examination of expanded concordances of shared c-collocates revealed further instances of conflation in the use of RASIM.
- Discourses related to RASIM seem to revolve around a small number of categories, most denoting negative stance.
- Overall, c-collocates index the treatment of RASIM as a problem.

## References

- Baker, P. (2004). Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics* 32(4): 346-359.
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, J.P. & McEnery, A.M. (2005). A corpus-based approach to discourses of refugees and asylum seekers. *Journal of Language and Politics*, 4(2), 197-226.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61-74.
- Gabrielatos, C. (2007). Selecting query terms to build a specialised corpus from a restricted-access database. *ICAME Journal*, 31, 5-43.
- Greenslade, R. (2005). Seeking scapegoats: The coverage of asylum in the UK press. Asylum and Migration Working Paper 5. Institute for Public Policy Research.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Koller, V. & Mautner, G. (2004). Computer applications in critical discourse analysis. In Coffin, C., Hewings, A. & O'Halloran, K. (eds.) *Applying English Grammar: Functional and corpus approaches*. London: Hodder and Stoughton, 216-28.
- Krishnamurthy, R. (1996). Ethnic, racial and tribal: The language of racism? In C-R. Caldas-Coulthard & M. Coulthard (eds.) *Texts and Practices: Readings in Critical Discourse Analysis*. London: Routledge, 129-149.
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Baker, M., Francis, G. & Tognini-Bonelli, E. (eds.) *Text and technology: In honour of John Sinclair*, Philadelphia and Amsterdam: John Benjamins, 157-176.
- McEnery, T. (2006). *Swearing in English: Bad language, purity and power from 1586 to the present*. London: Routledge.
- Orpin, D. (2005). Corpus linguistics and critical discourse analysis: Examining the ideology of sleaze. *International Journal of Corpus Linguistics*, 10(1), 37-61.
- Reisigl, M. & Wodak, R. (2001). *Discourse and Discrimination: Rhetorics of racism and anti-semitism*. London and New York: Routledge.
- Sedlak, M. (2000). You really do make an unrespectable foreign policy. In Wodak, R. & van Dijk, T.A. (eds.) *Racism at the Top: Parliamentary discourses on ethnic issues in six European states*, 107-168.
- Sinclair, J.McH. (1991) *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Scott, M. (1999). *WordSmith Tools Help Manual*. Oxford, UK: Mike Scott and Oxford University Press.
- Sinclair, J.McH. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sotillo, S.M. & Wang-Gempp, J. (2004). Using corpus linguistics to investigate class, ideology, and discursive practices in online political discussions: Pedagogical applications of corpora. In Connor, U. & Upton, T.A. (eds.) *Applied corpus linguistics*, 91-122.
- Scott, Mike. 2004. *Oxford WordSmith Tools Version 4*. Oxford: Oxford University Press. Available online: <http://www.lexically.net/downloads/version4/wordsmith.pdf>
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.
- ter Wal, J. (2002). Racism and cultural diversity in the mass media: An overview of research and examples of good practice in the EU Member States, 1995-2000. European Research Centre on Migration and Ethnic Relations (ERCOMER).
- van der Valk, I. (2000). Parliamentary discourse on immigration and nationality in France. In Wodak, R. & van Dijk, T.A. (eds.) *Racism at the Top: Parliamentary discourses on ethnic issues in six European states*, 221-260.
- van Leeuwen, Theo. (1996). The representation of social actors. In C-R. Caldas-Coulthard and M. Coulthard (eds.) *Texts and Practices. Readings in Critical Discourse Analysis*. London: Routledge, 32-70.
- Wodak, R. & van Dijk, T.A. (eds.) (2000). *Racism at the Top: Parliamentary discourses on ethnic issues in six European states*. Drava Verlag.