

## HEGEMONIC AND MINORITY DISCOURSES AROUND IMMIGRANTS: A CORPUS-BASED CRITICAL DISCOURSE ANALYSIS

Muchamad Sholakhuddin Al Fajri

Lancaster University

m.s.alfajri@gmail.com; s.alfajri@lancaster.ac.uk

First received: 25 February 2017

Final proof received: 22 September 2017

### Abstract

This study aims to analyse discourses surrounding the word *immigrants* in a large collection of naturally occurring language, 'ukWac' corpus (Web as Corpus). It employs corpus linguistics as a methodology to carry out critical discourse analysis research. Specifically, collocation analyses were used to identify dominant representations and discourse prosodies (Stubbs, 2007) of immigrants. Concordance analyses were then applied to examine the data in a more qualitative way. The findings suggest that while there are a few instances indicating positive representations of immigrants, hegemonic discourses around them are more negative. They are predominantly constructed as illegal entities, victims and dangerous groups. These constructions are likely to prime people to think that all immigrants are illegal and threatening, and will not be able to integrate into their host society.

**Keywords:** discourse analysis; critical discourse analysis; corpus linguistics; immigrants; construction

Discourse around immigrants has attracted ample attention from (critical) discourse analysts. Several studies have been conducted in different socio-political contexts, such as Flowerdew, Li and Tran's (2002) study on the representation of Chinese immigrants in Hong Kong, Teo's (2000) research on the discursive dimensions of the representation of immigrants in Australia and more recently Khosravinik's (2010) investigation into the discursive strategies used by numerous British newspapers to construct refugees, asylum seekers and immigrants. These previous studies, however, tend to employ various methods and approaches that are merely qualitative and focus more on news media discourse. Studying newspaper discourse can be an effective way to analyse dominant ideologies in a society (Bednarek & Caple, 2017). However, nowadays there are 'new media', such as blogs, online news and online forums, which can also affect public's perceptions. Therefore, it seems essential to study how 'new media' construct a particular group of people.

This study endeavours to combine quantitative and qualitative methods by using corpus linguistics as a methodology to conduct critical discourse analysis research. Also, instead of investigating news discourse, it analyses discourses surrounding the subject *immigrants* in the 'ukWac' corpus (Web as Corpus). To begin with, it is necessary to describe briefly the structure of this study. It begins with a literature review which respectively deals with the use of corpora in discourse analysis, some previous studies, several benefits of corpus-based approaches to discourse analysis, and some concerns. Data and method are then described before providing finding and discussion in the next section. Lastly, it presents the conclusions and limitations of this study.

### Corpus linguistics and discourse analysis

The integration of corpus linguistic techniques into discourse analysis has been shown as a fruitful methodology in language research (Baker et al., 2008). It seems that the synergy brings somewhat opposite approaches together. Discourse analysts generally use qualitative techniques and take several elements beyond texts such as social and political conditions into account, while corpus linguistic practitioners tend to employ quantitative methods or use statistical tests generated by computer software or web-based corpus analysis system, and examples of language use in a corpus are taken out of context. Therefore, some discourse analysts prefer not to use corpora. However, although corpus linguistics implements a more quantitative technique (such as employing frequency information), qualitative methods are often needed. As Biber, Conrad, and Reppen (1998: p. 4) points out, "Association patterns represent quantitative relations, measuring the extent to which features and variants are associated with contextual factors. However functional (qualitative) interpretation is also an essential step in any corpus-based analysis". In this way, corpus could be a valuable tool in discourse analysis in the sense that it can show repetitions and patterns of a particular linguistic phenomenon and can be used to identify implicit meanings behind the patterns and their influence upon people and on society.

A number of studies have applied corpus linguistic approaches to discourse analysis in order to provide insights into the relationship between language and society, ideology and power. Some linguists have used what Baker et al. (2008) called "a useful methodology synergy" to examine an ideology of words or a group of words, such as

Krishnamurthy (1996) who studied the words *ethnic*, *racial* and *tribal* and Orpin (2005) who analysed words semantically related to corruption (such as *sleaze*). It also has been used to examine a particular topic, such as Alexander's (1999) study of ecological issues in business texts and Baker's (2006) analysis on the issue of "fox-hunting" in British parliamentary debates, and topics more related to politics, such as Fairclough's (2000) analysis of New Labour's rhetoric and Mulderrig's (2011) investigation into New Labour's educational governance. Another application of corpus-based approaches to discourse analysis is to study the representation or construction of a particular group of people, which will be described in more detail in the following paragraphs since it seems more related to the present study.

Baker (2012) examines how Islam and Muslims are represented in the British press by using a corpus consisting of about 143 million words from 200.000 articles in the British national press between 1998 and 2009. He concludes that the British press are biased against Muslims since they have a tendency to over-focus on extreme Muslims who tend to be associated with very negative contexts such as terrorism and conflicts, or they frequently associate Islam with extremism.

Mautner (2007) analyses the stereotypical constructions of *elderly* in *Wordbanks Online* consisting of 57 million words. She found that *elderly* is mostly connected with discourses of "care", "disability", and "vulnerability", "emerging less as a marker of chronological age than of perceived social consequences" (p. 51).

More recently, in the case of analysing a corpus of news texts published in a specific news institution, Salahshour (2016) examines the portrayal of migrants in a 700.000-word corpus of news articles in *New Zealand Herald*, a daily New Zealand newspaper. The analysis suggests that liquid metaphors such as *influx(s)*, *inflow(s)* and *wave(s)* are not only used to represent mass immigration more negatively as suggested in the previous research but also in more positive ways, reflecting its positive impact on New Zealand economy.

The above-mentioned previous studies have shown that corpus-based approaches to discourse analysis can reveal a mainstream representation of a group of people or a particular social group by examining a large number of texts, which can also boost the empirical credence of the analysis. Corpus linguistic techniques that are mostly used in their studies are *collocation* and *concordance*. These techniques, therefore, are also employed in this present study.

The proponents of corpus-based approaches to discourse analysis state that this approach has several salient advantages. First and foremost, it will reduce "the cherry-picking" problem (Koller &

Mautner, 2004, p. 225; Partington, 2004, p. 13) or limit researcher bias (Baker, 2006). Discourse analysts are often criticised for only selecting data that suit their political agenda or support their initial hypothesis (Widdowson, 2004) and ignore those which show the contrary. For instance, we may only select a text or an article that confirms our perspective but ignore others, which present different viewpoints. By using a corpus, which stores numerous articles or texts, we can restrict our cognitive biases since "we are starting (hopefully) from a position whereby the data itself has not been selected in order to confirm existing conscious (or subconscious) biases" (Baker, 2006, p. 12).

The second benefit of using corpus-based approaches to discourse analysis is that it can show "the incremental effect of discourse" (Baker, 2006, p. 13). Discursive representations of events or people in society are not created through a single text or a single grammatical structure but they are constructed through repeatedly occurring patterns of language use. The cumulative effects of such patterns might shape discourses or the way people perceive the world. As Stubbs (2001, p. 215) contends, "Repeated patterns show that evaluative meanings are not merely personal and idiosyncratic, but widely shared in a discourse community. A word, phrase or construction may trigger a cultural stereotype." Therefore, corpus methods can be an ideal tool to demonstrate the cumulative effects since the application of corpora could provide a number of supporting instances of a discourse construction or show such repeated patterns.

Furthermore, this approach can enable us to spot "resistant and changing discourse" (Baker, 2006, p. 14). "Discourses are not static" (Ibid.) so that what was common-sense ways of viewing the world years ago could be different or unacceptable today. This change can be demonstrated by using diachronic (or historical) corpus since this type of corpus can be used to track changes in language evolution (McEnery, Xiao, & Tono, 2006; McEnery & Hardie, 2011), presumably including changes in discourses. Additionally, it is also conceivable to see the change by comparing more than one corpus which contains texts from different time periods.

Lastly, it could provide a safeguard against over-interpretation and/or under-interpretation in critical linguistic analysis of how a text positions its reader (O'Halloran & Coffin, 2004). Critical linguists often seek to examine ideological meanings in texts such as newspapers, tabloids, speeches etc. and therefore they read texts with a particular purpose which the casual readers may not have. In this way, critical linguists could overinterpret a text by only fixing values that they have noticed in the text and constructing meaning that may not be necessarily the same as the casual reader's perspectives. In contrast, it is also possible for critical linguists to under-interpret the text under

analysis if they are not a part of regular readers. Corpora can help reduce the prospect of over- and under interpretation by using the aid of a concordancer and general or specialized corpora. The use of corpora and concordancer can allow us to understand clearly “how meanings in a text are constructed to set up a dynamic reader position” (O’Halloran & Coffin, 2004, p. 294).

Although this method has considerable benefits, Baker (2006) in his book “Using Corpora in Discourse Analysis” has pointed out several concerns. This paper, however, only discusses two concerns that are arguably the most significant. First, discourses are not merely communicated through spoken and written language or verbal communication but they can also be conveyed through non-verbal communication, such as gestures and images or photographs. As Chouliaraki and Fairclough (1999, p. 38) point out,

“We shall use the term ‘discourse’ to refer to semiotic elements of social practices. Discourse therefore includes language (written and spoken and in combination with other semiotics, for example, with music in singing), nonverbal communication (facial expressions, body movements, gestures, etc.) and visual images (for instance, photographs, film). Indeed, spoken text can be considered multimodal, too, as it includes non-verbal modes of communication such as gesture, body posture and facial expression.”

Thus, since corpus data are mostly in the form of words (written or transcribed spoken), a corpus-based study might be restricted in the verbal domain (Baker, 2006) and cannot examine how different modes work together to construct discourses. In other words, it can be said that corpus-based approaches could not conduct multimodal discourse analysis.

Secondly, the fact that examples of language in a corpus do not tell anything about the context in which those examples were produced might become another concern (Baker, 2006) as discourse analysis generally put emphasis on the analysis of texts as well as interaction and social contexts (Fairclough, 2010). “We may not know the ideologies of the text producers in a corpus” (Baker, 2006, p. 18) since examples of language in corpora are decontextualized. However, it is argued that the social contexts can be identified through repeated patterns occurring in a corpus, and decontextualized data in corpora can also be valuable since it can encourage the researcher “to spell out the steps that lie between what is observed and the interpretation placed on those observations” (Hunston, 2002, p. 123).

## METHOD

The source of data in this present study is from ukWac corpus. The corpus is constructed by web-

crawling or automatically downloading texts from the web, limited to the .uk internet domain. It was created between 2005 and 2007 as part of the “WaCky project” (Web as Corpus kool ynitiative) and it consists of more than 2 billion tokens. According to Baroni, Bernardini, Ferraresi, & Zanchetta (2009, p. 209), “ukWaC is among the largest, and the only English web-crawled resource with linguistic annotation”, and it is aimed to “serve as general-purpose resources for English”. One of the salient advantages of using this type of corpus might be that it contains genres that are not found in traditional written corpora such as blogs and online discussion forums. This study only uses 50% randomly selected sample of ukWac corpus, which contains 1,127,056,026 words. In addition to 50% sample of ukWac corpus, a more general English corpus, i.e. British National Corpus (henceforth BNC), was employed for reference in some circumstances. The BNC corpus, however, is not applied directly to analyse discourses of immigrants but it is referred to, since “it can reveal normative patterns of language use which can then be compared against the findings” in the ukWac corpus (Baker & McEnery, 2005, p. 200).

There are various corpus linguistic methods that can be applied to discourse analysis. This study mainly employs collocation techniques as collocations can “convey messages implicitly and even be at odds with an overt statement” (Hunston, 2002, p. 109), which might reveal a number of discourses. As Baker (2006) points out, using a collocational analysis can be valuable in discourse analysis since it can provide the most salient and clear lexical patterns surrounding a subject, from which discourses could be acquired. Also, because this study uses a large corpus containing more than ‘1 billion words’, a collocational analysis can help focus on initial analysis and avoid from sorting concordances multiple times to uncover lexical patterns (Baker, 2006). In addition to collocation techniques, concordance analysis is also used as a supplement in order to avoid over or under interpretation of collocation data. The concordance analysis focuses on adjectival, noun and verbal collocates, as well as “semantic prosody” or as Stubbs (2007, p. 178) prefers to call it “discourse prosody” since it is a topic where corpus linguists and (critical) discourse analysis practitioners have mutual interest (Koller & Mautner, 2004, p. 222). Louw (1993: 157) defines the term “semantic prosody” as “[a] consistent aura of meaning with which a form is imbued by its collocates”. One of Louw’s (p. 160) examples in the phenomenon of “semantic prosody” is the word “utterly”. According to his analysis, the word “utterly” has an “overwhelmingly” negative prosody as it often collocates with words that have negative meanings such as “exhaustive”, “ridiculous”, “terrified” (Ibid.).

In conducting collocational analysis, it is significant to be clear about a collocation span, cut-off points and statistical measures to calculate collocations since a different span, cut-off points and measures will result in different collocates. This paper uses a span of three words on either side of the node word. In terms of statistical measures, Log-Likelihood (Dunning, 1993) was used because it measures collocations by significance, the higher the score indicating the association is not by chance, and it does not give high scores to fairly low frequency words as the Mutual Information (Berry-Rogghe, 1973) measure does (see McEnery et al., 2006). A search for the word “immigrants” returned 9,451 matches in 5,048 different texts and it has 9,687 collocate words. However, I only considered 15 best adjectival, verbal and noun collocates that have high log-likelihood scores and at least occur 5 times in the corpus.

## FINDINGS AND DISCUSSION

The first stage of the analysis involves obtaining collocations of *immigrants* by using online corpus software CQPweb (Hardie, 2012). It has several available corpora including the 50% sample of ukWac corpus and can allow us to analyse concordances, collocations, frequency lists, keywords, distribution tables and charts (Ibid.). The analysis of collocations in this study focuses only on the adjectival, noun and verbal collocates since they seem to reveal much interest, particularly in

discourse analysis. Thus, function words or grammatical words such as conjunctions, articles, prepositions are removed from the collocate lists.

### Adjectival Collocates of Immigrants

Following Sinclair’s (2003, p. xvi) statement, “Decide on the strongest pattern and start there”, the collocate which has the highest log-likelihood score was first examined. From Table 1, it can be seen that the word *immigrants* is strongly associated with the adjective *illegal*. It collocates with immigrants 1,031 times in the corpus. It seems that the combination of *illegal* with *immigrants* indicates a negative discourse prosody in itself since the word “illegal” carries a negative semantic meaning. When we have a close look at the concordances in Table 2, we also can see that the combination is mostly related to negative situations. In line 1, “illegal immigrants don’t work”. Line 2 refers to “illegal immigrants feared drowned”. Line 3 suggests that immigrants place a heavy financial burden on the government and in line 4 illegal immigrants are detained. Moreover, the combination is linked by the conjunction “and” with apparently more serious problems, suggesting that they are in the same group. For example, in line 5, immigrants are linked with smuggling and drugs. Lines 6 and 7 show that immigrants are connected with terrorism. In line 8, immigrants are linked with criminals, drugs, plant and animal diseases and rabies. These negative constructions may emphasise that immigrants are a part of enormous problems in society.

Table 1. Adjectival collocates of *immigrants*

No	Word	Observed collocate frequency	Log-likelihood
1.	Illegal	1,031	11437.455
2.	Legal	185	767.228
3.	Irish	120	665.835
4.	Jewish	103	632.267
5.	New	393	555.207
6.	Undocumented	45	550.003
7.	Many	282	532.463
8.	Nonelderly	26	426.494
9.	Chinese	73	360.666
10.	Recent	112	336.572
11.	Asian	58	323.589
12.	Would-be	31	264.971
13.	Italian	55	258.474
14.	Muslim	49	245.638
15.	medicaid	27	228.065

Besides the word *illegal*, *immigrants* also collocates with the word “undocumented”, which seems to have less negative implications, 45 times in the corpus. Although *illegal* and *undocumented* appear to have a similar meaning, the word “illegal” carries more negative connotations since it commonly collocates with an inanimate entity such as drugs, logging etc. Indeed, if we look at collocates of *illegal* in the BNC, we will find that the top 10 collocates that have high log-likelihood

scores are mostly non-living objects, such as *drugs*, *activities*, *trade*, *abortion*, *logging*, *rave* and *drug*. Thus, the combination of *illegal* and *immigrants* seems to more criminalise the people rather than their actions, and it can also dehumanise them.

It is interesting to note that the word *immigrants* also collocates with the adjective *legal* 185 times in the corpus, which has the second highest log-likelihood score. However, the fact that there is a wide gap between the number of the word *legal*

(185) with the word *illegal* (1,031) that co-occur with *immigrants* will bring about a negative representation of immigrants in society. As Stubbs (1996, p. 195) points out “if collocations and fixed phrases are repeatedly used as unanalysed units in

media discussion and elsewhere, then it is very plausible that people will come to think about things in such terms”. Therefore, this construction might lead people to think that all immigrants are illegal and immigration is illegitimate.

Table 2. Concordance of immigrants when it collocates with *illegal*

No	Concordances
1.	2:32 pm I thought that the Mail was always complaining about <b>illegal</b> Immigrants who <b>do n't work</b> . Now they 're working , and they
2.	as floodwaters rose . 14th A group of up to 163 <b>illegal</b> Immigrants feared drowned while trying to reach Australia from Indonesia likely sailed into
3.	result , Taiwan 's detention centers are frequently jammed with <b>illegal</b> Chinese immigrants , which places a heavy financial burden on the government . The
4.	Mail ( 29.09.98 ) reported that five were detained as suspected <b>illegal</b> Immigrants , one was charged with a firearms offence and the seventh man
5.	ways of tackling the everpresent problems of smuggling , drugs and <b>illegal</b> Immigrants . The course also covers the various charterparties affecting the agent ,
6.	raised by those who saw the proposal as an avenue for <b>illegal</b> immigrants and terrorism . There were sufficient safeguards to guarantee against these dangers
7.	are : to increase security , both against potential terrorists and <b>illegal</b> Immigrants ; to expedite travel for legitimate travellers ; to create efficiencies in
8.	criminals , drugs , plant and animal diseases and rabies and <b>illegal</b> immigrants " . Despite the fears of Mrs Thatcher , the Single Market

Within the best 15 adjectival collocates that have high log-likelihood scores, 4 collocates refer to a name of a country or a continent and religion, suggesting that immigrants are also frequently described in terms of where they are from (Irish, Chinese, Asia and Italian) and what religion they belong to (Jewish and Muslim).

Another adjectival collocate that might be significant to be considered is the word “would-be”. It collocates 31 times with immigrants. Would-be immigrants also tend to have a negative prosody. From table 3, we can see this pattern. In line 1, would-be immigrants are *unwelcome* in a particular country, in line 2 they *died* in the sea, in line 3 and 6 would-be immigrants are also *illegal* and line 4 shows that they had been *detained*. Additionally, some concordances indicate that they are victims or in danger. Line 5, for instance, states that controls on immigrants are harsh and brutal, and line 6 says that they are in vulnerable position. On top of that,

line 7 uses the word “invasion” which is metaphorically employed to construct immigrants as dangerous nuisances because “invasion” is strongly related to an occasion when a country is taken or controlled forcedly by another country, and to a process or an act that makes someone’s life unpleasant. Therefore, these patterns could also sustain negative representations of immigrants since the people who still want to be immigrants already have associated with treacherous situations or actions and are represented as a dangerous group for their host country.

The other adjectival collocates (*new*, *many*, *nonelderly*, *recent* and *medicaid*) might not be extremely essential to be analysed deeply. Only might the word “medicaid” need to be briefly explained. *Medicaid* occurs mostly in the phrase *immigrants’ medicaid*, meaning health insurance coverage of immigrants.

Table 3. Concordance of immigrants when it collocates with *would-be*

No	Concordances
1.	government would be launching a propaganda offensive in China to inform <b>would-be</b> Immigrants of how unwelcome they are here . And he suggested that a
2.	The Association for Human Rights in Andalucia has claimed that 289 <b>would-be</b> immigrants died in the sea between Spain and Morocco last year , up
3.	14 months - From Ananova - A total of 55 <b>would-be</b> illegal immigrants have been prevented from entering the UK this weekend by a ferry
4.	maritime rescue authorities said they had detained a total of 196 <b>would-be</b> Immigrants after four makeshift vessels tried to make it to the shore on
5.	programmes are a major cause of emigration . Controls imposed on <b>would-be</b> Immigrants are harsh and often brutal, as governments aim to stop the
6.	will promote prevention, i.e. will send discouraging signals to <b>would-be</b> illegal Immigrants and those who exploit their vulnerable positions ,and make the European
7.	now seamlessly integrated into British life . The present invasion by <b>would-be</b> Immigrants presents a different problem . Many of them speak little or no

**Noun collocates of immigrants**

The top 2 noun collocates which have the highest log-likelihood scores are “seekers” and “asylum” (see table 4 for the noun collocate list). Both words refer to a single phrase “asylum seekers” (the fourteenth collocate in table 4). Thus, they will be analysed together as one phrase. Additionally, the word *immigrants* also strongly collocates with the word “refugees” (the third collocate in table 4). The phrase *asylum seekers* and the word *refugees* mainly co-occur with immigrants because they are connected by the conjunction “and”, indicating that they are also in the same group. According to Baker and McEnery’s (2005) analysis of the representation of asylum seekers and refugees in the UN and newspaper texts, both asylum seekers and refugees possess a negative meaning and they are constructed as “a problematized group” (p. 216). For example, the word “asylum” is also associated “with mental

illness and incarceration” (Baker & McEnery, 2005, p. 214) and “refugees are constructed as a ‘natural disaster’ like a flood” (p. 204). Thus, although immigrants, asylum seekers and refugees have a distinct meaning that might carry different international obligations and consequences, grouping them together can highlight the negative representation of immigrants. Moreover, besides *asylum seekers* and *refugees*, *immigrants* are commonly linked together by the conjunction “and” with the word “minorities”, emphasising that they are also part of a minority group.

Furthermore, the word *immigrants* commonly collocates with the word “numbers”, pre-modifying quantification. This pattern occurs 85 times in the corpus. In some cases, this quantification indicates that the volume of immigrants is problematic (see table 5 for examples).

Table 4. Singular and plural noun collocates of *immigrants*

No	Word	Observed collocate frequency	Log-likelihood
1.	seekers	192	1753.880
2.	asylum	195	1538.147
3.	refugees	170	1429.938
4.	influx	70	730.972
5.	Post	149	494.141
6.	minorities	54	419.439
7.	natives	40	361.868
8.	filipino	31	330.390
9.	immigrants	45	320.701
10.	descendants	38	311.416
11.	insurance	104	290.775
12.	prwora	28	285.832
13.	numbers	85	275.512
14.	asylum-seekers	25	270.783
15.	nonreplacement	16	248.596

Table 5. Concordance of immigrants when it collocates with *numbers*

No	Concordances
1. Black and Asian presence in Britain ) " The presence of coloured	immigrants in large <b>numbers</b> resulted in serious racial <b>tensions</b> which the country had
2. throughout the Highlands . There is always <b>tension</b> when large <b>numbers</b> of	immigrants arrive in a new area and start to displace the indigenous population
3. In areas where the indigenous population feel <b>threatened</b> by large <b>numbers</b> of	immigrants , groups such as Combat18 and the BNP are growing rapidly
4. love , fear , happiness and sadness of the enormous <b>numbers</b> of	immigrants <b>flooding</b> into Argentina who were trying to start a new life at
5. complexly than census forms allow . However classified , record <b>numbers</b> of	immigrants are <b>streaming</b> to the United States and creating the most diverse society

In line 1 and 2, the large numbers of immigrants caused tension and in example 3 the large numbers of immigrants made “the indigenous population feel threatened”. Additionally, in some circumstances this combination (*numbers* and *immigrants*) is also connected to some movement metaphors that compare immigrants to water in some way (see line 4 and 5 in table 5). This is congruent with Baker and McEnery’s (2005) analysis of *refugees*. In this sense, then immigrants are also dehumanised and represented as disaster like a “flood”, “which is

difficult to control as it has no sense of its own agency” (p. 204).

The last noun collocate that is significant is the word “descendants”, which co-occurs with immigrants 38 times. This pattern might highlight that immigration has occurred for several generations. Moreover, in the BNC, the word *descendants* significantly collocates with the preposition “of” which is mostly followed by proper nouns and mainly a name of famous or important people. Thus, it seems that this pattern has a more

positive portrayal and it is also likely to suggest that immigrants are essential part of society.

The other noun collocates seem not critically important in this study. *Influx* nearly always occurs in the phrase “influx of immigrants”, meaning arrivals of immigrants in a large number. *Filipino* might be wrongly tagged since it is an adjective. *Insurance* mainly occurs in the phrase “health insurance” – insurance for immigrants.

**Verbal Collocates of Immigrants**

Table 6 shows the top 15 verbal collocates of *immigrants*. The verbal collocate that significantly associates with the word *immigrants* is the verb *integrate*. Its literal meaning is similar to the collocate *assimilate* and therefore they will be

examined together as one group. In some cases, this pattern is used to construct immigrants as subjects of help (see line 1, 2, 3, 4 in table 7), which might indicate that they cannot integrate to their host society by themselves effortlessly. In fact, taking a closer look at the concordances appears to confirm this assumption. The concordances in table 7 suggest that the integration of immigrants into their host society needs great effort and it is often unsuccessful and failed. This tends to emphasise the view that immigrants are powerless in integration processes and could also implicitly say that immigrants are less educated since a common way to help immigrants integrate into society is by learning the language of their host country or by going to school.

Table 6. Verbal (present) collocates of *immigrants*

No	Word	Observed collocate frequency	Log-likelihood
1.	integrate	18	83.782
2.	investigate	22	71.297
3.	deport	8	69.257
4.	settle	17	68.753
5.	assimilate	7	51.236
6.	fix	14	46.090
7.	blame	11	35.345
8.	attract	12	31.639
9.	fill	11	15.844
10.	arrive	9	15.228
11.	accept	14	14.454
12.	identify	13	9.902
13.	Post	22	7.959
14.	stop	15	7.600
15.	solve	5	6.580

Table 7. Concordance of immigrants when it collocates with *integrate* and *assimilate*

No	Concordances
1. Immigrant Absorption - operates in a number of ways to <b>help</b> new	immigrants <b>integrate</b> into the job market in accordance with their education and work
2. we short of in Britain today ? How can we <b>help</b> new	immigrants to <b>integrate</b> ? Irma Reid , who came from Trinidad to work
3. also from Vietnam , Indonesia , and Thailand . To <b>help</b> these	immigrants <b>assimilate</b> into Taiwan society , the government has earmarked NT\$3 billion (
4. discrimination helping young people grow up with a sense of belonging <b>helping</b>	immigrants <b>integrate</b> into our communities Find out more about our plans to achieve
5. Turkey 's accession is Europe 's experience of the <b>failure</b> of Moslem	immigrants to <b>integrate</b> and share in mainstream European democratic traditions and values
6. ( 4.45 billion euros ) <b>unless serious steps</b> are taken to <b>integrate</b>	immigrants better. 4 September 2003 ( 96 words ) Government plans curbs
7. The other fear was of l'inassimilablité? , the concern that these	immigrants would <b>not integrate successfully</b> into French society
8. had benefited Britain. And almost two-thirds of whites said they believe	immigrants <b>do not integrate</b> or make a positive contribution to Britain. On
9. France , of course -- France <b>does n't know how to assimilate</b>	immigrants , and , as the BBC reported on Nov. 5 , 2005

The term *blame* occurs 11 times in the corpus in reference to immigrants (see table 8 for the concordance lists). Line 4 shows that blaming immigrants is somewhat inevitable. Line 6 and 7 seem to indicate a positive meaning but the word

“all” might imply that some/most immigrants are deserved to be blamed. However, it is important to note that several concordance lines apply distancing techniques. Line 1, for example, suggests that the “BNP blames immigrants”, line 2

talks about “the idea” that immigrants are to blame, line 3 states that something is “trying” to blame immigrants and line 9 talks about “the attempt” to

blame immigrants. Therefore, this could be more discourse which casts immigrants as victims.

Table 8. Concordance of immigrants when it collocates with *blame*

No	Concordances
1. n't afford a house and the BNP uses this issue to <b>blame</b>	immigrants . " The part rent , part buy housing which is pushed
2. aimed at controlling immigration , was designed to promote the idea that	immigrants were <b>to blame for unemployment</b> , poor housing and diseases . Of
3. promised during the election . Worse , they are trying to <b>blame</b>	immigrants from Eastern Europe ! Local Lib Dem campaigner Catherine Reifen was not
4. as a whole from it . I 'm not one to <b>blame</b>	immigrants for everything <b>but the fact that</b> they are working here could end up being a very big problem if the economy does go tits
5. your horizons become brighter . Of course , you must <b>blame</b> the	immigrants <b>for everything</b> that goes wrong in the British society and in your
6. because we need many scientists and entrepreneurs . One should appreciate genuine	immigrants and not <b>blame</b> them <b>all</b> . zzg113 Dec 27 2004 , 12:33 PM
7. You then go on to say " One should appreciate genuine	immigrants and not <b>blame</b> them <b>all</b> . " - Yes , I absolutely
8. finding that epidemics are not only another crisis to <b>blame</b> on	immigrants , but can also serve as an appropriately xenophobic metaphor for immigration
9. miraculously benefiting from free health care . The attempt to <b>blame</b>	immigrants for the NHS crisis has also led to the formation of a

The next verbal collocate to consider is the word “fill”, which co-occurs 13 times with immigrants. Table 9 demonstrates 4 examples of the concordances. This combination seems to bring a more positive discourse prosody. For instances, in line 1, immigrants fill a shortfall, in line 2 immigrants fill vacancies quickly, in line 4 immigrants “helped to fill the gaps in national labour markets”.

The last collocate that needs to be examined is the word “solve”. On the surface, this collocate may suggest that immigrants are a problem that needs to be solved. However, the concordances reveal that instead of being a problem, immigrants are described as a solution to several problems (see table 10 for 4 examples of concordances). Nevertheless, since these patterns only have 6.58 log-likelihood score and only occur 5 times, this tends to be a minority discourse of immigrants.

Table 9. Concordance of immigrants when it collocates with *fill*

No	Concordances
1. once part of the British Empire . Historical sources indicated that many	immigrants came to <b>fill a shortfall</b> in labour in Britain during the post-war
2. 2004 were " upbeat about the outcome of this decision " .	Immigrants <b>helped to fill the gaps in national labour markets</b> and enlargement "
3. April 6 Home Secretary David Blunkett said there was a need for	immigrants to help <b>fill 60,000 " hospitality and catering " vacancies in London</b>
4. maintain the illusion of keeping by bringing in ever greater numbers of	immigrants to <b>fill low paid jobs</b> in the public and private sector and

Table 10. Concordance of immigrants when it collocates with *solve*

No	Concordances
1. introducing temporary immigration status to selected	immigrants to <b>solve short-term worker shortages</b> and reduce political tensions
2. but were.cThat make the arrived legal	Immigrants <b>solve the problem the voluntary schip</b>
3. businesses argue that they need a sizeable flow of	immigrants to <b>solve labour and skills shortages</b>
4. Somewhat likely to arrived legal	immigrants <b>solve the problem in designing institutions</b>

## CONCLUSION

This present study has attempted to examine discourses of immigrants. Overall, the results suggest that there are different discourses surrounding the word. On the one hand, it might be

stated that this study has revealed common representations or ‘hegemonic discourses’ of immigrants. They are mainly represented as ‘illegal’ entities which have a more series of negative implications. This representation also dehumanises

them since the strongest collocates of *illegal* are inanimate entities. Moreover, they are commonly described as powerless in integrating to their host country, as victims and dangerous nuisances. These representations are likely to prime people to think that all immigrants are illegal and threatening, and will not be able to integrate into their host society. On the other hand, in few cases, immigrants are also constructed positively. In other words, they still have a few positive associations, whereas the mainstream discourses of immigrants are more negative. For example, they are sometimes represented as a solution to economic problems, such as filling the gaps in national labour markets and solving labour and skills shortages, but it is a minority discourse.

It should be noted, however, that this study still has some limitations. The fact that it only considers the top 15 collocates that have high log-likelihood scores and analyses verbal collocates only in its present form can conceal some interesting discourses. Therefore, further study could consider more collocates which may uncover different patterns of discourses. Additionally, the application of other corpus software which has a more range of features might also be recommended.

#### ACKNOWLEDGMENTS

I would like to express my deep gratitude to Paul Baker for his valuable feedback on the earlier version of this manuscript, and to the anonymous reviewers for their constructive comments. I am also immensely grateful to Indonesia Endowment Fund for Education (LPDP) for funding this study.

#### REFERENCES

- Alexander, R. J. (1999). Ecological commitment in business: A computer-corpus-based critical discourse analysis. In J. Verschueren (Ed.), *Language and ideology: Selected papers from the 6th International Pragmatics Conference* (pp. 14-24). Antwerp: International Pragmatics Association.
- Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Baker, P. (2012). Acceptable bias? Using corpus linguistics methods with critical discourse analysis. *Critical Discourse Studies*, 9(3), 247-256. doi:10.1080/17405904.2012.688297
- Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, M., McEnery, A. M., & 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), 273-306.
- Baker, P., & McEnery, T. (2005). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics*, 4(2), 197-226.
- Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources & Evaluation*, 43(3), 209-226. doi:10.1007/s10579-009-9081-4
- Bednarek, M., & Caple, H. (2017). *The discourse of news values: How news organizations create newsworthiness*. New York: Oxford University Press.
- Berry-Rogghe, G. L. E. (1973). The computation of collocations and their relevance in lexical studies. In A. J. Aitken, R. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 103-122). Edinburgh: Edinburgh University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. New York: Cambridge University Press.
- Chouliaraki, L., & Fairclough, N. (1999). *Discourse in late modernity: rethinking critical discourse analysis*. Edinburgh: Edinburgh University Press.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61-74.
- Fairclough, N. (2000). *New labour, new language?* New York: Routledge.
- Fairclough, N. (2010). *Critical discourse analysis: The critical study of language* (2nd Ed.). Harlow: Longman.
- Flowerdew, J., Li, D. C. S., & Tran, S. (2002). Discriminatory news discourse: Some Hong Kong data. *Discourse & Society*, 13(3), 319-345.
- Hardie, A. (2012). CQPweb: Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380-409. doi:10.1075/ijcl.17.3.04har
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Khosravinik, M. (2010). The representation of refugees, asylum seekers and immigrants in British newspapers: A critical discourse analysis. *Journal of Language and Politics*, 9(1), 1-28.
- Koller, V., & Mautner, G. (2004). Computer applications in critical discourse analysis. In C. Coffin, A. Hewings, & K. O'Halloran (Eds.), *Applying English Grammar* (pp. 216-228). London: Arnold.
- 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* (pp. 129-149). London: Routledge.

- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, E. Tognini-Bonelli, & J. M. Sinclair (Eds.), *Text and Technology* (pp. 157-176). Amsterdam: John Benjamins.
- Mautner, G. (2007). Mining large corpora for social information: the case of elderly. *Language in Society*, 36, 51-72.
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.
- Mulderrig, J. (2011). Manufacturing consent: A corpus-based critical discourse analysis of New Labour's educational governance. *Educational Philosophy and Theory*, 43(6), 562-578. doi:10.1111/j.1469-5812.2010.00723.x
- O'Halloran, K., & Coffin, C. (2004). Checking overinterpretation and underinterpretation: Help from corpora in critical linguistics. In C. Coffin, A. Hewings, & K. O'Halloran (Eds.), *Applying English grammar* (pp. 275-297). London: Arnold.
- Orpin, D. (2005). Corpus linguistics and critical discourse analysis: Examining the ideology of sleaze. *International Journal of Corpus Linguistics*, 10(1), 37-61. doi:10.1075/ijcl.10.1.03orp
- Partington, A. (2004). Corpora and discourse, a most congruous beast. In A. Partington, J. Morley, & L. Haarman (Eds.), *Corpora and Discourse* (pp. 11-20). Bern: Peter Lang.
- Salahshour, N. (2016). Liquid metaphors as positive evaluations: A corpus-assisted discourse analysis of the representation of migrants in a daily New Zealand newspaper. *Discourse, Context & Media*, 13, 73-81. doi:10.1016/j.dcm.2016.07.002
- Sinclair, J. M. (2003). *Reading concordances: An introduction*. Harlow: Longman.
- Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. London: Blackwell.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. London: Blackwell Publishers.
- Stubbs, M. (2007). Quantitative data on multi-word sequences in English: The case of the word world. In M. Hoey (Ed.), *Text, discourse and corpora: Theory and analysis* (pp. 163-189). London: Continuum.
- Teo, P. (2000). Racism in the News: A critical discourse analysis of news reporting in two Australian newspapers. *Discourse & Society*, 11(1), 7-49.
- Widdowson, H. G. (2004). *Text, context, pretext: critical issues in discourse analysis*. Oxford: Blackwell.