Media representations of ‘leftover women’ in China: a corpus-assisted critical discourse analysis

Yu Yating

Abstract

The term ‘leftover women’, commonly referring to single women older than 27, has been in popular use in Chinese media since 2007. This study investigates how leftover women are linguistically represented in the English-language news media in China by employing a corpus-assisted approach to critical discourse analysis. A specialised corpus of 303 English news articles (i.e. 236,254 words), covering the years between 2007 and 2017, was built for this purpose. Corpus linguistics techniques were employed to quantify the meaning shift units (MSUs) of the lemma leftover WOMAN, and van Leeuwen’s ‘social actors and actions theory’ was applied to inform the classification of MSUs in context. These findings shed light on media representations of leftover women, the contested ideologies emerging from these representations, and how shifting gender politics and identity shapes and is shaped by media in the world’s most populous nation.

KEYWORDS: LEFTOVER WOMEN; MEDIA REPRESENTATIONS; CORPUS-ASSISTED CRITICAL DISCOURSE ANALYSIS; GENDER IDEOLOGIES

The origins and meanings of leftover women

The term leftover women (Chinese character: 剩女; Pinyin: shèngnǚ) was first introduced in the media by the All-China Women’s Federation in the year 2007 (Fincher 2014), and it was officially listed as one of the neolo-
gisms by the Ministry of Education of the People’s Republic of China in the same year (Li and Li 2013). The term ‘leftover woman’ is defined as an ‘unmarried female, usually older than one who is expected to get married’ (Li and Li 2013:307). The marriageable age is commonly regarded as no more than 27 (Fincher 2014).

A decade later, the All-China Women’s Federation published a news article urging the banning of a number of terms that they called ‘sexist’ including leftover women (H. Wang and Yang 2017), reflecting a rapid change in their (and the state’s) attitude to this term that they had once coined. Although it is not easy to speculate on the motivations behind this, the change might be relevant to the public attention demonstrated by the growing number of commercial and academic publications related to the leftover women phenomenon over the past decade (Chen 2013; Feldshuh 2018; Fincher 2014, 2016; Gaetano 2009, 2014; Ji 2015; Lake 2018; A. Luo, Wang and Jiang 2014; To 2013, 2015a, 2015b; You, Yi and Chen 2016). Some of these publications indicate that the concept of leftover women was deliberately created to pressure women into early marriage so as to achieve the government’s demographic goal of solving the declining birth-rate crisis, the ageing population problem, and China’s gender imbalance (Fincher 2014, 2016). These negative commentaries may have contributed to this rapid change of official attitudes towards the term.

**Representations of unmarried women**

*Representation*, as defined by Hall (1997), is a meaning-making process through language. Representations do not merely reflect or mirror reality but serve to represent or to construct the reality and identities of a social group from particular ideological perspectives (Gauntlett 2008). Similarly, Fairclough (1995) suggests that choices are made (e.g. what to include or exclude, background or foreground, and what lexis and grammar to use) in representations depending on the producers’ goals, knowledge, interest, shared values and the beliefs of their media organisations and the wider community.

In the West, linguistic representations of single women have attracted researchers’ attention, such as Romaine’s (2000) and Baker’s (2008a) investigation of ‘spinster’. Both of these studies compare the collocations of the lemma SPINSTER with BACHELOR in the British National Corpus of 100 million words.² Their findings indicate that BACHELOR and its collocates tend to have positive discourse prosody depicting unmarried males as desirable whereas SPINSTER and its collocates are more likely to have negative prosody indicating unmarried females are undesirable and miserable.
Similarly, in China, media discourses on single women have been explored by a few studies through content analysis (A. Luo et al. 2014), critical discourse analysis (W. Luo and Sun 2015) and narrative analysis (Feldshuh 2018). However, these studies mainly investigated the local Chinese-language media in China.

Luo, Wang and Jiang (2014) employed content analysis (simply by counting the frequency of contents) to investigate the representation of leftover women in 592 Chinese-language news texts produced by the Chinese Media between 2004 and 2011. Their findings indicate that the majority of the news articles (92.5%) focus on leftover women’s relationships and sexuality, but not so much on their career and other life skills. The media perpetuated the stereotype of leftover women as a group of self-centred, independent and ambitious women aged between 28 and 36 possessing high socio-economic status. Although they long for marriage, they have high expectations of their potential partners.

Luo and Sun (2015) examined a popular speed-dating televised show in China, named *If You Are the One*, which thrives on the leftover women phenomenon. They deployed CDA as a critical lens examining the use of metaphors, labels and insinuations in several aspects of the show such as the female candidates’ stage positions, the female candidates’ self-descriptions, the male and female candidates’ descriptions of their ideal partners, and the host and commentators’ narratives. Their study shed light on the construction of gender ideologies which reinforce the perceptions of ideal femininity as submissive, attractive and youthful in the televised show.

Similarly, Feldshuh (2018) investigated three Chinese televised programs: *If You Are the One* (a speed-dating reality show), *Let’s Get Married* (a soap opera) and *iPartment* (a situation comedy) focusing on the construction of the ‘leftover women myth’. She deployed narrative analysis attending to the aspects of humour, indirect sexism and cultural references which were intentionally used as social policing to preserve gender norms and hierarchy. The results show that the narratives in the three case studies constructed a uniform myth that a woman’s happiness lies in her marriage which might conflict with her pursuits, such as career and higher education.

Their findings coincidentally pinpoint the traditional patriarchal ideologies that the Chinese-language media discourses surrounding single women in China lean towards. Past traditional gender roles of women can be best depicted through the following Confucian proverbs (Bell and Chaibong 2003): the virtue of a woman lies in her obedience to the father, husband and son; men are in charge of the outside whereas women are in charge of the inside; the virtue of a woman lies in her lack of talent. Hence, the traditional cultural norms of femininity expected a woman to be dependent,
Yu Yating

submissive and unambitious. Women were obliged to produce offspring and served as the carers of the households. Therefore, wives sometimes were called ‘néirén’ (inside people) by their husbands. In ancient times, Confucian values had advantages in promoting harmonious families in which both men and women had clear-cut gender roles (Sung and Pascall 2014). Under the predominant influence of Confucian culture, the traditional practice of heterosexual marriage and harmonious families still prevails in Chinese society (Sung and Pascall 2014). As many traditional Confucian values and patriarchal norms are deeply rooted in Chinese society, not being able to get married might violate these values and norms and consequently put pressure on leftover women. China’s current political direction seems to be promoting traditional gender values more strongly recently and there has been a surge of support for Confucian ideals (Fincher 2014, 2016).

**Synergy between critical discourse analysis and corpus linguistics**

*Critical discourse analysis* (CDA) is a transdisciplinary field primarily concerned with the relationships between language use and social structure with its main aim to study how social power, domination and inequality are enacted, maintained and reproduced by texts (Fairclough 2010, 2015). Discourse is the favoured vehicle of ideologies. Ideologies are the common beliefs and shared values of a group. Ideologies reside in discourse and discourse reflects the ideologies of a certain group. They serve to exercise and sustain power to define or unite a group. Although dominant discourses are naturalised by ideologies, a critical scrutiny of the discourses can help to make them transparent. What the change in support and usage of the term leftover women might reflect about the Chinese government’s ideology is therefore valuable to study.

*Corpus linguistics* (CL) is a methodology, which employs computational techniques to study a large body of empirical data. One of the fundamental concepts in CL is that language is a social phenomenon; language and culture are inseparable. Therefore, linguistic evidence can be collected for analysing society. Displaying the concordance lines, comparing the repeated patterns and counting the frequencies could help to reveal relevant socio-cultural patterns.

Discourse studies that employ the synergetic methodology between CL and CDA are termed corpus-assisted CDA. CDA is qualitative in nature, and it has been criticised for subjectively cherry-picking data and interpreting data based on researchers’ own agendas (Widdowson 2000). On the other hand, CL has also been criticised for lacking in-depth analyses of contextual elements due to its descriptive and quantitative nature.
Corpus-assisted CDA has been proven useful in overcoming some of these respective issues and enabling broad analysis to be conducted in depth (Baker 2008b; Cheng 2013; Mautner 2009; O’Halloran 2010). For instance, to reduce researcher bias, CL techniques can be utilised to systematically uncover linguistic patterns and tendencies on a large scale (Cheng 2013; Gabrielatos and Duguid 2015). In addition, to defuse the ‘cherry-picking’ charges in CDA, CL provides different downsizing techniques to objectively select prototypical or event-based sample texts (Anthony and Baker 2015; Baker, Gabrielatos, Khosravinik, McNerney and Wodak 2008). In response to the criticism of CL lacking in-depth analyses, CDA approaches enable researchers to examine and interpret contextual features in both expanded concordances and full texts if necessary.

This methodological synergy is a ‘relatively new movement in linguistics’ and ‘the number of such studies in proportion to the number of studies in CL or CDA is extremely small’ (Baker et al. 2008:274). Although the visibility of corpus-assisted CDA has increased in different domains, including gender and sexuality, over the past decade (Nartey and Mwinlaaru 2019), only a few studies are found in the literature employing corpus-assisted CDA to investigate gender representations, such as Caldas-Coulthard and Moon (2010), Baker (2014) and Al-Hejin (2015).

Caldas-Coulthard and Moon (2010) examined how males and females were represented ideologically in broadsheets and tabloids by analysing the adjectival collocates preceding the singular gendered nouns (i.e. boy, girl, man and woman) and classifying them into different sociosemantic categories (van Leeuwen 1996) in the BoE corpus. Similarly, Baker (2014) conducted a diachronic study to investigate gender stereotypes and the usage of gender-biased language, which was informed by van Leeuwen’s (2008) sociosemantic approach in the 400-million-word Corpus of Historical American English (COHA). Likewise, Al-Hejin (2015) combined the discourse-historical approach (Reisigl and Wodak 2001), the sociosemantic approach (van Leeuwen 1996) and the sociocognitive approach (van Dijk 2008) to interrogate the referential and predicational strategies surrounding HIJAB in concordances representing Muslim women in the 1.9-million-word BBC news corpus.

Differing from the previous studies, this study applies both Sinclair’s (1996, 2004) notion of meaning shift units from CL and van Leeuwen’s (2008) sociosemantic approach from CDS, combining descriptive (quantitative) analysis with contextual (qualitative) analysis, to examine representations of leftover women and contribute to the growing literature of corpus-assisted CDA in the domain of gender representation.
The present study

This novel study is a pioneering work that examines linguistic representations of leftover women in the English-language news texts produced by the Chinese Media. These news articles were originally written in English (i.e. not translations of Chinese-language news) mainly by Chinese journalists.

This paper also explores whether the traditional gender ideologies are similarly portrayed in the Chinese English-language news texts; or if there are other emerging contested ideologies since the English-language news media, such as *Global Times*, claim that they aim to report ‘hard-hitting facts, controversial angles and daring argument’ that are free of political conspiracy, distinguishing themselves from other local Chinese-language media (*Global Times* 2018).

Finally investigating the English-language news articles could shed light on how the topic of leftover women is presented to the global audience illustrating one of the contemporary gender issues in modern China as well as perhaps shifts in gender concepts. With these objectives in mind, this study seeks to answer the following research questions:

1. How are leftover women linguistically represented in the English-language news texts produced by the Chinese media?
2. What ideologies are portrayed from the linguistic patterns of representations?
3. How do these ideologies illustrate, support and conflict with traditional and emerging concepts of gender?

Data collection

In order to build an English corpus of media representations of leftover women (MRLWC) covering the period between 2007 and 2017, relevant data were collected by searching the phrase ‘leftover woman/women’ and its synonyms, namely ‘leftover lady/ladies’ and ‘leftover girl(s)’ from two sources: *China Daily* online news website and WiseNews archive. The WiseNews archive contains Mainland Chinese newspapers, such as *Shanghai Daily, Global Times* and *Shenzhen Daily*. The type of news in the corpus belongs to ‘soft news’, which includes features on lifestyle, business, society, culture, regional stories and opinion. During data collection, each article was checked manually to ensure that the primary topic of the news is relevant and related to leftover women.

The news content is stored in plain text files since standard concordancers require semiotically reduced texts, which are ‘stripped of formatting,
layout and accompanying visuals’ (Mautner 2009:129). The metadata of the news is stored in a separate Excel file that contains the information on topics, authors, newspapers, word count, web addresses, publishing and collection dates so that the original files and information can be retrieved. This metadata can inform the analysis of discursive practice by providing the information on authors and newspapers and track webpage layouts if multimodality is concerned.

To check and remove duplicated texts, WordSmith v. 6.0 (Scott 2013) was employed to compute a twelve-word cluster list with a minimum frequency of two times. 46 articles were found to contain identical clusters. However, it should be noted that duplicate clusters do not necessarily mean duplication of whole texts (Baker 2014). After manually reading through the paragraphs where the identical clusters lie, 17 duplicate articles were identified and removed. The remaining 29 articles were retained as their identical clusters are quotations, proper nouns, or clauses. In addition, shen nu was substituted for sheng nu, shengnü, sheng nü, shengnv, shengnyu, sheng nyu, shengnü1 and shengnu1 as these are interchangeable written forms of Pinyin (the romanisation system of Mandarin Chinese) for leftover women found in the corpus. In the end, the corpus contains 303 news articles with 236,254 words.

Analytical framework and procedures

Sinclair’s (1996, 2004) notion of meaning shift units (MSUs; also called lexical items or extend units of meaning) was employed to identify and classify the associated meanings of leftover WOMAN (329), which has been proven to be the canonical form of the translation of shèngnǚ based on its frequency compared with leftover LADY (13) and leftover GIRL (12). The concept of MSUs was first proposed by Sinclair in 1996 as ‘a higher rank of lexical structure, above the word’, which functions as a basic meaningful set through the co-selection of different words (Sinclair 2004:25). MSUs consist of five types of co-selection. The definition of each is indicated in Table 1.

Among the five elements of MSUs, semantic prosody, also named discourse prosody by Stubbs (2001), is considered as contentious (Hunston 2007), as varied scholars use this term differently either as specific discourse functions (Sinclair 1996, 2004; Stubbs 2001) or distinct attitudes (i.e. negative and positive) (Partington 2004). To clarify the definition, this study regards semantic prosody as functional meanings, the interpretations relying on the immediate co-text of the MSUs.
There are two main analytical steps. Firstly, in order to investigate the collocation and colligation of *leftover* WOMAN and their co-occurring patterns, the Concord function in WordSmith v. 6.0 was employed to generate collocates of *leftover* WOMAN (calculating both *leftover woman* and *leftover women* together with the wildcard asterisk * sign) within a span of five words on either side of the search term. As a wide span might elicit irrelevant collocates whereas a narrow span might overlook non-contiguous collocates, WordSmith’s default setting was selected as a compromise to avoid these issues. The minimum frequency for collocates to be considered is eight times, which resulted in 59 collocates. The reason for choosing a frequency-based measure over other statistical ones (e.g. log-likelihood and \( T \)-score) to obtain the collocates is that language data (which is constrained by paradigmatic, syntagmatic and other elements) do not occur randomly (Kilgarriff 2005; Stubbs 1995); as Baker (2014:135) indicates the issue with employing the statistical measures is that ‘they are based on comparing what has actually occurred in a corpus against a model of words occurring in a completely random order across a corpus’. Admittedly, the choices made for the collocational span and the minimum frequency of occurrence is subjective, but it resulted in a manageable (in terms of time constraints and word limits) set of collocates, accounting for 70 per cent of total recurrent collocates in this study. After fixing the set of collocates, they were categorised into function words and content words. Although some words are polysemous with more than one word class (e.g. *Chinese* as a noun and adjective), the classification is based on their most frequent senses co-occurring with *leftover* WOMAN.

Secondly, ConcGram v. 1.0 (Greaves 2009) was employed to investigate the semantic preference and semantic prosody of *leftover* WOMAN. Complementing the functions in WordSmith v. 6.0, ConcgGram v. 1.0 is able to search for all the ‘configuration’ of a lexical set, including the ‘positional and constituency variation’ (Cheng, Greaves, Sinclair and Warren 2009;
Cheng, Greaves and Warren 2006:414; Cheng and Lam 2010, 2013), such as leftover women, leftover single women, women are leftover. ConcGram v. 1.0 generated concordance lines with three-word concgrams – that is leftover + WOMAN + its frequent co-selected content words as the origins. Although 50 characters (i.e. about 12 words) on either side of the centred is the maximum span by default, ConcGram v. 1.0 enables access to full articles from the concordance lines for more context when necessary. It is worth noting that the wildcard ‘%’ was applied in ‘wom%n’ to ensure that both singular and plural forms were included in the search. In the end, the search resulted in 682 concordance lines.

These 682 concordance lines were exported to an Excel spreadsheet for qualitative analysis (i.e. manual identification and classification) as not all leftover WOMAN and its frequent co-selected content words in the 682 concordance lines constitute MSUs. For instance, in considering the association between leftover women and its collocate China in the following two examples:

leftover women’ in China

What is it like being a leftover woman who wants to be a mother? In China, there...

The first example is identified as an MSU because China is associated with leftover women as a post-modifier whereas the second one would not be considered as an MSU because leftover women and China do not associate with each other. In addition, concordance lines that contain leftover MAN were excluded from the analysis. In the end, a total of 292 MSUs were identified.

Furthermore, these 292 MSUs were manually analysed in context (i.e. by referring back to the full articles of the MSUs) and classified into different semantic preferences and semantic prosodies, informed by van Leeuwen’s (2008) social actor and action theory, so as to uncover the ideologies from the linguistic patterns of representations.

These steps employed the collocation and concordance tools to quantify the linguistic patterns of MSUs which were further subject to qualitative interpretation and critical scrutiny so as to unveil the possible ideologies behind them. In addition, stepping outside the corpus (when necessary) to consult wider contexts, such as a dictionary and a reference corpus, can inform the interpretation and explanation of the linguistic patterns. Through investigating linguistic patterns, CDA can be a starting point to understand better other dimensions of society, namely, discursive and sociocultural practice (Fairclough 2010, 2015). These findings are discussed in the next section.
Analysis: the invariable core

At the initial stage of analysis, consulting a dictionary and reference corpus could provide additional contextual information about the term leftover women as an invariable core.

The adjective leftover is defined as ‘remaining after all the rest has been used, taken, or eaten’ by the Longman Dictionary of Contemporary English. In addition, searching the word leftover in the British National Corpus (BNC) returned 55 hits shown in concordance lines. These results indicate that leftover tends to be used as an adjective (44 out of 55) rather than a noun in the BNC corpus. Although there is one creative usage where leftover is used to describe humans (And people, leftover people like – like grubs that’ve been kept in the dark), the adjective most commonly modifies food (59.1%), other commodities and objects (e.g. paint) (25%), and abstract concepts (e.g. love) (13.6%). Therefore, by modifying women with the adjective leftover, the phrase leftover women implies an analogy that this kind of women are like food, commodities, or objects that are left behind after the rest have been taken or consumed.

The Chinese phrase shèngnǚ can be alternatively translated as leftover GIRL or leftover LADY. Table 2 indicates that leftover WOMAN (329) produced the highest frequency compared with leftover LADY (13) and leftover GIRL (12). Sinclair (2005) proposes a few figures for considering the frequency of a searched term: two is the minimum frequency for a term to be regarded as recurrent and twenty is the minimum occurrence for a term to be examined without ambiguities by a lexicographer. Therefore, this study focuses on the phrase leftover WOMAN as the canonical form of the translation of shèngnǚ based on its prominent frequency, and excludes the other two terms from analysis as their frequencies are too low for making any generalisations of their linguistic patterns. Although the frequency of leftover WOMAN (329) is relatively low, it is still adequate for studying this single item in contexts based on the minimum frequency (i.e. 20 times) suggested by Sinclair (2005).

Table 2: Total instances of leftover WOMAN, GIRL and LADY.

<table>
<thead>
<tr>
<th>Lemmas</th>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>leftover WOMAN</td>
<td>329</td>
</tr>
<tr>
<td>leftover LADY</td>
<td>13</td>
</tr>
<tr>
<td>leftover GIRL</td>
<td>12</td>
</tr>
</tbody>
</table>
In addition to their usage being too low for in-depth study, it is worth noting that the other two terms, GIRL and LADY, rather than WOMAN in the English translation also carry different implications. GIRL can refer to either female children or young women according to the Longman Dictionary of Contemporary English. However, previous corpus studies have shown that GIRL tends to refer to female adults associated with gendered or sexualised representations (Baker 2014; Caldas-Coulthard and Moon 2010; Norberg 2016; Sigley and Holmes 2002; Taylor 2013). As for LADY, previous corpus research (Holmes 2000; Holmes and Sigley 2001) indicates that it is regarded as a ‘pseudo-polite’ word which has been ‘seen as patronising, conservative and dated’ and its occurrence has been decreasing (Sigley and Holmes 2002:150).

**Collocations and colligations**

Table 3 shows the frequent collocates of leftover WOMAN (ranked by frequency and reverse alphabetical order) and Table 4 shows the word classes of the collocates. The top ten frequent collocates for both forms are mostly (i.e. nine out of ten) function words. Among the top 10 collocates, the function words at L1 (i.e. one word to the left) are *a, about, and, are, as, being, for, like, many, most, not, of, on, or, over, than, that, the, to and with*; at R1 (i.e. one word to the right), the function words are *a, and, are, as, but, by, for, has, have, I, in, is, like, many, more, of, or, she, that, the, to, was, who and with*. The indefinite article *a* tends to appear at L1 (62%) of leftover woman and at R2 (30%) of leftover women, which distinguishes between the singular and plural forms – the former individualises a social actor (e.g. *even my parents call me a leftover woman now*) whereas the latter indicates a neologism (e.g. ‘leftover women’ is a derogatory but wildly used term).

The majority of the content words (55%) are nouns. The co-selected nouns represent leftover WOMAN either as a neologism or phenomenon. **Phrase** (38%) and **term** (39%), which regard leftover WOMAN as a neologism, tend to appear at L1 (e.g. *the phrase ‘leftover women’ and the term ‘leftover women’*). **Women** (40%) tends to appear at R2 or R5, classifying what kinds of women are termed ‘leftover’ (e.g. ‘leftover women’ unmarried women of a certain age and ‘leftover women’ or educated, urban, professional women). **Shengnu** (42%), the Chinese translation of leftover WOMAN, tends to appear at L2 immediately followed by the adverb **literally** (38%) at L1 (e.g. *shengnu, literally ‘leftover women’*). **Phenomenon** (71%), which represents leftover WOMAN as an intangible social phenomenon, tends to appear at R1 (e.g. *the ‘leftover women’ phenomenon*). China’s (70%), tends to appear at L1 as a premodifier (e.g.
Table 3: Frequent collocates of *leftover* WOMAN.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Word</th>
<th>Freq.</th>
<th>Rank</th>
<th>Word</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>THE</td>
<td>142</td>
<td>31</td>
<td>OVER</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>108</td>
<td>32</td>
<td>THAN</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>OF</td>
<td>99</td>
<td>33</td>
<td>ABOUT</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>TO</td>
<td>90</td>
<td>34</td>
<td>WAS</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>AND</td>
<td>75</td>
<td>35</td>
<td>ON</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>SHENGNU</td>
<td>63</td>
<td>36</td>
<td>I</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>IN</td>
<td>59</td>
<td>37</td>
<td>LABELLED</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>IS</td>
<td>55</td>
<td>38</td>
<td>BY</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>OR</td>
<td>54</td>
<td>39</td>
<td>MANY</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>AS</td>
<td>49</td>
<td>40</td>
<td>WITH</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>ARE</td>
<td>44</td>
<td>41</td>
<td>BECOME</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>TERM</td>
<td>39</td>
<td>42</td>
<td>MORE</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>WOMEN</td>
<td>34</td>
<td>43</td>
<td>BEEN</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>THAT</td>
<td>32</td>
<td>44</td>
<td>NOT</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>MEN</td>
<td>32</td>
<td>45</td>
<td>NUMBER</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>FOR</td>
<td>30</td>
<td>46</td>
<td>BUT</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td>WHO</td>
<td>25</td>
<td>47</td>
<td>MARRIAGE</td>
<td>10</td>
</tr>
<tr>
<td>18</td>
<td>CALLED</td>
<td>21</td>
<td>48</td>
<td>SHE</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>BEING</td>
<td>20</td>
<td>49</td>
<td>LIKE</td>
<td>9</td>
</tr>
<tr>
<td>20</td>
<td>LEFTOVER</td>
<td>20</td>
<td>50</td>
<td>MOST</td>
<td>9</td>
</tr>
<tr>
<td>21</td>
<td>BE</td>
<td>19</td>
<td>51</td>
<td>THEIR</td>
<td>8</td>
</tr>
<tr>
<td>22</td>
<td>SINGLE</td>
<td>19</td>
<td>52</td>
<td>SUCH</td>
<td>8</td>
</tr>
<tr>
<td>23</td>
<td>HAS</td>
<td>18</td>
<td>53</td>
<td>WHICH</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>CHINA</td>
<td>18</td>
<td>54</td>
<td>THEY</td>
<td>8</td>
</tr>
<tr>
<td>25</td>
<td>HAVE</td>
<td>18</td>
<td>55</td>
<td>ONE</td>
<td>8</td>
</tr>
<tr>
<td>26</td>
<td>CHINA’S</td>
<td>16</td>
<td>56</td>
<td>LITERALLY</td>
<td>8</td>
</tr>
<tr>
<td>27</td>
<td>SO</td>
<td>15</td>
<td>57</td>
<td>PRESSURE</td>
<td>8</td>
</tr>
<tr>
<td>28</td>
<td>PHENOMENON</td>
<td>14</td>
<td>58</td>
<td>PHRASE</td>
<td>8</td>
</tr>
<tr>
<td>29</td>
<td>SAID</td>
<td>14</td>
<td>59</td>
<td>AT</td>
<td>8</td>
</tr>
<tr>
<td>30</td>
<td>UNMARRIED</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
China’s ‘leftover women’) whereas China (42%) tends to appear at R2 as a postmodifier (e.g. ‘leftover women’ in China), indicating the geographical location. Number (80%) tends to appear at L2 (e.g. the number of ‘leftover women’) aggregating all the leftover women with statistics. Men (25%) tends to appear at R2 (e.g. leftover women and men) assimilating single men and women as one homogeneous group. Marriage (20%) tends to appear at R3 associating leftover women with marriage (e.g. ‘leftover women’ in the marriage market).

Table 4: Linguistic property of frequent collocates.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Word class</th>
<th>Collocates of leftover WOMAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function words</td>
<td>Auxiliary verbs</td>
<td>are, be, been, being, has, have, is, was</td>
</tr>
<tr>
<td></td>
<td>Conjunctions</td>
<td>and, but, or</td>
</tr>
<tr>
<td></td>
<td>Determiners</td>
<td>a, many, more, most, that, the</td>
</tr>
<tr>
<td></td>
<td>Intensifiers</td>
<td>so</td>
</tr>
<tr>
<td></td>
<td>Negative marker</td>
<td>not</td>
</tr>
<tr>
<td></td>
<td>Prepositions</td>
<td>about, as, at, by, for, in, like, of, on, over, than, to, with</td>
</tr>
<tr>
<td></td>
<td>Pronouns</td>
<td>her, I, one, she, such, their, they, who</td>
</tr>
<tr>
<td>Content Words</td>
<td>Nouns</td>
<td>China, China’s, marriage, men, number, phenomenon, phrase, Shengnu, term, women</td>
</tr>
<tr>
<td></td>
<td>Verbs</td>
<td>become, called, labelled, pressure, said</td>
</tr>
<tr>
<td></td>
<td>Adjectives</td>
<td>leftover, single, unmarried</td>
</tr>
<tr>
<td></td>
<td>Adverbs</td>
<td>literally</td>
</tr>
</tbody>
</table>

The co-selected adjectives (i.e. leftover, single and unmarried) describe the marital status of leftover WOMAN. Leftover (40%) tends to appear at L3 (e.g. leftover men and women). Single (21%) tends to appear at R1 (e.g. ‘leftover women’, single women in China). Unmarried (15%) tends to associate meaningfully with leftover WOMAN at R1 (e.g. ‘leftover women’ unmarried women of a certain age).
The co-selected verbs are the actions of *leftover WOMAN* or others. *Become* (45%) tends to appear at R2 (e.g. *'leftover women' has become the subject matter*). *Called* (57%) tends to co-occur at L1 (e.g. *called 'leftover women'*). *Labelled* (67%) tends to appear at L2 or L3 (e.g. *labeled as 'leftover women' or labeled as a 'leftover women'*. *Said* (29%) tends to appear at R1 (e.g. *leftover women, said Chen*). *Pressure* (38%) tends to appear at R3 as a nominalised action (e.g. *'leftover women' has mounted pressure*).

**Semantic preferences and prosodies**

Three semantic preferences, informed by van Leeuwen's (2008) social actors and actions theory, are identified, namely:

1. ‘neologism’ (e.g. *'leftover woman’, a term that refers to women who remain single after thirty in China*);
2. ‘action’ by others and leftover women (e.g. *I was suddenly labeled as a leftover woman by my parents and society and they would grow old and become one of the many “leftover women”*); and
3. ‘social phenomenon’ (e.g. *US journalist Roseann Lake has investigated the phenomenon of ‘leftover’ women in China*).

Each semantic preference is associated with different semantic prosodies (see Table 5). The following sections discuss each category respectively in detail.

**Neologism**

The semantic preference of ‘neologism’ (57.5%) defines *leftover WOMAN* with the most frequent co-selected nouns (i.e. *phrase, Shengnu, term and women*), adjectives (i.e. *single and unmarried*) and adverbs (i.e. *literally*). The definitions are given immediately before or after where *leftover WOMAN* appears (Extract 1). The definitions explain what *leftover WOMAN* is regarding their age, education, professions, independence, appearance, salary, personal success, marital status and geographic locations. Identifiable ages for *leftover WOMAN* found in the concordance lines are 27, 28, 29, 30, 32 and 35. Among these ages, 27 co-occurs most frequently (i.e. six times) with *leftover WOMAN* and is recognised as the starting point for women who are *deemed too old for marriage* (Extract 2). They have high social economic status (i.e. good education, high income and good occupations) and they are highly independent and live in the urban areas of China. Their appearance is considered attractive, but their marital status is *single or unmarried. Leftover WOMAN* as a neologism
often requires explanation when appearing in the news; this could be because the target readership is English-speaking readers worldwide including visiting professionals and tourists who might not be familiar with this apparently culture specific term. In addition to the definitions, the neologism’s inventors are also mentioned (e.g. the All-China Women’s Federation in Extract 3).

**Extract 1**

The Ministry of Education listed shengnu, literally ‘leftover women’ (but actually single, especially educated and well-employed, women) as a new word in 2007. Such women set high standards for the men they would marry, and, in the process, many of them end up staying single beyond their traditional marriage age.

---

**Table 5:** Semantic preferences and semantic prosodies of leftover WOMAN (LW).

<table>
<thead>
<tr>
<th>Semantic preferences</th>
<th>Semantic prosodies</th>
<th>Freq. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neologism (n = 168; 57.5%)</td>
<td>otherness</td>
<td>83 (49.4)</td>
</tr>
<tr>
<td></td>
<td>neutrality</td>
<td>41 (24.4)</td>
</tr>
<tr>
<td></td>
<td>disapproval of neologism</td>
<td>36 (21.4)</td>
</tr>
<tr>
<td></td>
<td>distinguished women</td>
<td>8 (4.8)</td>
</tr>
<tr>
<td>Action (n = 73; 25%)</td>
<td>social discrimination</td>
<td>44 (60.3)</td>
</tr>
<tr>
<td></td>
<td>passiveness</td>
<td>14 (19.2)</td>
</tr>
<tr>
<td></td>
<td>credibility</td>
<td>13 (17.8)</td>
</tr>
<tr>
<td></td>
<td>neutrality</td>
<td>2 (2.7)</td>
</tr>
<tr>
<td>Social phenomenon (n = 51; 17.5%)</td>
<td>neutrality</td>
<td>29 (56.8)</td>
</tr>
<tr>
<td></td>
<td>social controversy</td>
<td>14 (27.5)</td>
</tr>
<tr>
<td></td>
<td>uniqueness</td>
<td>5 (9.8)</td>
</tr>
<tr>
<td></td>
<td>emancipation</td>
<td>2 (3.9)</td>
</tr>
<tr>
<td></td>
<td>profitability</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>292 (100)</td>
</tr>
</tbody>
</table>

---
Extract 2

Women over 27 are shengnu, ‘leftover women’ deemed too old for marriage.

Extract 3

I don’t think that women’s status today is higher than in the olden days. The oppressions that Chinese women suffer at present are more covert than before: for instance, the derogatory term ‘leftover women’ was created to pressure unmarried women to marry and give birth. Ironically, this insulting terminology was backed in 2007 by the All-China Women’s Federation.

The semantic prosodies of ‘otherness’, ‘neutrality’, ‘disapproval of neologism’ and ‘distinguished women’ are identified when the semantic preference of ‘neologism’ occurs. The neologism tends to generate the meanings of otherness (49.4%). In this case, the neologism classifies single women older than 27 as ‘others’, especially those spending their youth on climbing the social ladder instead of finding a husband. For instance, female PhD students, whom are frequently (74 times) mentioned in the whole corpus, are categorised into the third category of humankind (Extract 4). It implies that heterosexuality and marriage is the norm (i.e. heteronormativity; Warner 1993) and those not following these practices are seen as deviant or queer in Chinese society (Cameron and Kulick 2003).

Extract 4

Her father is worried that she could become a ‘third category person.’ ‘He had been surfing the Internet and read a post that says there are three categories of people: men, women and female PhD students,’ Xie said. ‘He was so afraid that the degree might make me unsuitable for marriage.’

Leftover’ status is often attributed to women’s increasingly high standards for men. Accusations of having unrealistically high expectations of potential partners are manifested in the attributive adjective picky, the action spend a long time looking for an ideal husband in Extract 5, and set high standards for the men they would marry in Extract 1.
Extract 5

During the recession, ‘affordable men’ have replaced rich bankers and successful managers as the first choice for women, especially those ‘leftover women’ (pretty, picky women who spend a long time looking for an ideal husband but end up single in their 30s), who want to settle down.

On the other hand, 21.4 per cent of the semantic prosodies disapprove the use of this term describing it as scornful, horrible, derogatory and so on (Extract 3). Regarded as a form of oppression, this term generates fear among women and pressures them into marriage (Extract 3). There are eight instances (4.8%) where the neologism is used positively, describing these women as distinguished, such as being independently successful (Extract 6). When the neologism is used neutrally (24.4%), it tends to introduce the phrase purely as data, such as reporting a survey (Extract 7).

Extract 6

Now that gold-diggers are made into the envy of young girls, shouldn’t we reserve some respect for women who want to be seen as independently successful? This brings us to ‘leftover women’, a term that applies to many career women.

Extract 7

The survey also explored respondents’ attitudes about shengnu, which literally means ‘leftover women’.

Action

The majority (76.7%) of actions are enacted by other participants of society whereas the rest (23.3%) are enacted by leftover women. The frequent co-selected material actions (represented as actual ‘doing’) by others are called, labelled and pressure. Called and pressure are interactive involving two participants, and leftover WOMAN is the one being called and pressured by others. Labelled is instrumental (actions that have an effect on another human or nonhuman goal/subject) and metaphorical, classifying women into categories as if they are nonhuman objects. In addition, called and labelled are in the passive voice (e.g. is often called), mostly (97%) leaving the agents omitted (Extract 8). Similarly, pressure is either nominalised (e.g. the pressure for in Extract 9) or served as a verb in the infinitive form (e.g. to pressure in Extract 10), sometimes (63%) omitting
the agents. The *exclusion* of agents do not show who is responsible for the discriminatory acts of calling, labelling and pressuring; as van Leeuwen (2008:30) indicates that ‘the point is that the practice is here represented as something not to be further examined or contested’. On the other hand, there are four occasions where the agents are identified as families, society and state media (Extracts 11 and 12). The semantic prosodies of actions by others tend to be associated with discriminatory meanings (60.3%) and in these cases, leftover women are subject to labelling, name calling and pressure. When the semantic prosody is neutral (2.7%), the actions co-occur with leftover women without taking any stance, such as naming a book (Extract 13).

Extract 8

A woman who isn’t married, doesn’t have a boyfriend and is over 30 years old is often called a ‘leftover woman’ in China.

Extract 9

The pressure for ‘leftover women’ in cities to marry has pushed them to go to extremes.

Extract 10

The oppressions that Chinese women suffer at present are more covert than before: for instance, the derogatory term ‘leftover women’ was created to pressure unmarried women to marry and give birth.

Extract 11

‘Now I am in a PhD program, and somehow I was suddenly labeled as a leftover woman by my parents and society.’

Extract 12

The subsequent promotion in State media of the ‘leftover’ label is intended to put pressure on women to hurry up and get married and bear children, she says.

Extract 13

Leta Hong Fincher, who has written a book called *Leftover Women: The Resurgence of Gender Inequality in China.*
When leftover WOMAN is the agent (10.9%), their frequent co-selected material actions are non-transactive (actions involving only one participant – the actor) and affective (reactions involving feelings). Become is non-transactive which involves only the actor and does not affect others other than themselves (Extract 14). Van Leeuwen (2008:61) indicates that ‘the actions of lower-status actors are more often represented as non-transactive’. Pressure is affective, representing the way how leftover women react to their single status. ‘As the power of social actors decreases, the amount of emotive reactions attributed to them increases’ (van Leeuwen 2008:56). In addition, pressure is always deactivated through nominalisation, treating their reaction as an entity rather than feelings (Extract 15). In contrast with the actions of others which are represented as actual ‘doing’, leftover women are engaged with passive reactions to their ‘leftover’ status. They do not voluntarily choose to be ‘leftover’, but unwillingly become leftover. Therefore, the semantic prosody of actions by leftover women is associated with ‘passiveness’ (19.2%).

Extract 14

They can’t walk away from the men because they live in the perpetual fear that they would grow old and become one of the many ‘leftover women’.

Extract 15

Single Chinese women face the pressure of becoming ‘leftover’.

When the action is semiotic (represented as saying or meaning), the frequent co-selected verb said is used to quote the leftover women and others who were interviewed. In the case of others, the interviewees are identified as authors, experts, business people, parents and wife seekers who shared their opinions towards leftover women and the phenomenon. In the case of leftover women, even though they are given a voice, it is always negative. They revealed their fear of becoming ‘leftover’ (Extract 16). These interviewees are nominated and individualised by presenting their surnames or full names and some (especially experts) by titles and institutional affiliations to show their credibility and authority (e.g. Hao Pengfei, director of the committee). The intertextuality of the texts (i.e. the frequent use of reporting verbs said) has reflected the practice of journalism to present news with credibility through direct and indirect quotes from the sources; therefore, in this case, the semantic prosody is associated with the meanings of ‘credibility’ (17.8%).
Extract 16

‘My parents have been nagging me to find a husband since I turned 25. As times goes by, I panic at times about becoming a shengnu (leftover woman)’, Li said.

Social phenomenon

When the semantic preference of leftover WOMAN is ‘social phenomenon’ (17.5%), it represents leftover WOMAN as an abstract and intangible social phenomenon. In this case, leftover WOMAN is mainly used in the plural form, and its frequently co-occurring words are China, China’s, number and phenomenon. When phenomenon co-occurs with leftover WOMAN, it refers to leftover WOMAN as a social phenomenon and the frequent collocates China and China’s indicate the provenance of the distinct social phenomenon (Extract 17). The collocate number is used to assimilate or aggregate leftover WOMAN as a homogeneous social group with statistics (Extract 18).

Extract 17

US journalist Roseann Lake has investigated the phenomenon of ‘leftover’ women in China.

Extract 18

A survey found that the number of shengnu, or ‘leftover women’, is rising in Shanghai because women are believed to be too aggressive in relationships, according to the Shanghai newspaper Labor Daily.

When the semantic preference of ‘social phenomenon’ occurs, the majority (56.8%) of semantic prosodies are neutral without any stance, such as media coverage of the leftover women phenomenon (Extract 17). 27.5 per cent of the semantic prosodies regard the leftover women phenomenon as controversial. For instance, it indicates that the logic behind the phenomenon is twisted especially when there is a surplus of men in China (Extract 19). 9.8 per cent of the semantic prosodies describe the leftover women phenomenon as unique and culture-specific (Extract 20). 3.9 per cent of the semantic prosodies regard the phenomenon as a form of emancipation for Chinese women who are no longer confined to the domestic domain; instead, they are playing a significant role in the public domain (Extract 21). 2 per cent of the semantic prosodies regard the phenomenon as a profitable opportunity for business (Extract 22).
Extract 19

How can a woman be too old to find a spouse in such a gender-skewed society? Yet, the ‘leftover women’ phenomenon has its own twisted logic.

Extract 20

Here we have to mention the unique leftover women or shengnu phenomenon in China.

Extract 21

From a social perspective, perhaps this is a contributing factor as to why there are so many ‘leftover women’ in China, a popular term to describe adult females who have chosen careers over marriage, embracing a new social revolution of independence, career and individual choice.

Extract 22

Wu said the ‘leftover women’ phenomenon will fuel market demand for her program, as well as other services.

Unveiling the ideologies of leftover women

Four distinct ideologies can be interpreted and summarised from the findings, namely ageism, heteronormativity, patriarchy and egalitarianism. First, age is crucial in classifying these unmarried women regarding their marriageability. The ageist discourse shows that 27 is a significant turning point for a woman’s value in the marriage market; especially those who spend their prime time on climbing the social ladder instead of finding a husband might risk becoming unwanted like leftover food. One of the reasons might be due to the biological fact that women’s fertility declines with age. Another reason might be due to the traditional Chinese concept of ideal femininity as youthful and attractive.

In addition, patriarchal ideology prevails in the representations. High socioeconomic status is always highlighted in the classification of leftover women because it subverts the traditional cultural norms of femininity as dependant, submissive and unambitious. In modern China, the traditional concepts of hypergamy (women ‘marrying up’) and ‘men dang hu dui’ (marriage between two families of equal socioeconomic status) still prevail (Ji 2015; To 2015b; J. Wang 2011). Therefore, the patriarchal discourse perpetuates the myth that the higher up a woman climbs the social ladder, the
more difficult it is for her to find a partner. In this culture a man may feel uncomfortable that his status is weakened if he were to marry a woman of higher economic rank.

Moreover, the ideology of heteronormativity underlies the representations of leftover women. Women older than 27 without being married are seen as ‘other’ in a society where heterosexual and early marriage practice is the norm (Ji 2015; J. Wang 2011). Therefore, they are constantly labelled and discriminated against by other social members. The media also pushes women to lower their standards so as not to fall into this unmarried ‘other’ group.

Meanwhile, discrimination against unmarried women is contested. Instead of pitying their singleness and childlessness, the egalitarian discourse regards the leftover women phenomenon as a form of emancipation and empowerment for Chinese women as their roles and values are redefined in the process of marketisation and social transformation. They are no longer confined by wifely duties and motherhood in the private sphere; they are equally as capable and distinguished as men serving in the public sphere.

Conclusion

This study has demonstrated how leftover women are linguistically represented and explored the ideologies in the Chinese English-language news media by employing a corpus-assisted approach to critical discourse analysis. Leftover women are found to be associated with neologism, a social phenomenon, and various actions which underlie the linguistic representations.

The representations found in this study conform to the stereotypes of leftover women as single women older than 27 associated with ‘three highs – high education, high professional achievement and high income’ and perpetuate the traditional ideology of women’s roles and values depicted in the local Chinese-language news media (Fincher 2014:15; A. Luo et al. 2014). Meanwhile, the English-language news discourse is more contentious as conservative (i.e. ageism, heteronormativity, patriarchy) and progressive (i.e. egalitarianism) gendered ideologies co-exist.

Admittedly, the frequency of leftover WOMAN found in this corpus is relatively low. Further studies could consider collecting English-language news about leftover women from UK or US sources (e.g. the WebCorp tool and GloWbE corpus), which will offer insights into different gender ideologies between the West and the East.

In addition, the representations of leftover women have reflected a social phenomenon that is also concerned with leftover men, due to the ageing
population, gender imbalance and the growing trend towards comparatively late marriage. Conducting an in-depth analysis of the comparison between leftover women and leftover men in the future will shed light on gender differences and similarities.

The way forward, hopefully, will witness the use of such a derogatory term – leftover women – falling into a decline in the media. Although the term was banned from using by the All-China Women’s Federation in the year 2017, to what extent such a ban is effective in reality remains a question. Only a future diachronic corpus-assisted study could inform whether the frequency of the term has decreased in the media since 2017.

About the author

Yu Yating is a PhD candidate from the Department of English, The Hong Kong Polytechnic University. Her research interests are in gender studies, corpus linguistics and (critical) discourse analysis. Her paper ‘Gender and conversational humor in a televised situational comedy: implications for EFL contexts’ appeared in English for Specific Purposes World in 2014.

Acknowledgements

I would like to thank Phoenix Lam, Malcolm MacDonald, Daniel Barlow and two anonymous reviewers for their very insightful comments on an earlier version of this paper.

Notes

1 The All-China Women’s Federation is a state-sponsored non-profit organisation, which is set up to protect women’s rights.
2 Lemmas, written in capital letters, include all the related grammatical forms. For example, WOMAN includes its singular and plural forms (i.e. woman and women).
5 The Lancaster BNCweb Server is available at http://bncweb.lancs.ac.uk/bncweb.

References


Partington, A. (2004). ‘Utterly content in each others company’: semantic prosody and


