LEMMATIZING TEXTBOOK CORPUS FOR LEARNER DICTIONARY OF BASIC VOCABULARY

Gusti Astika
Satya Wacana Christian University, Indonesia
gustiastika@yahoo.com

First received: 11 September 2017 Final proof received: 30 January 2018

Abstract
The purpose of this study is to lemmatize the vocabulary used in textbooks with the purpose to provide information for dictionary entries that are relevant to the learners. Although dictionaries for learners are commercially and electronically available, they may not be relevant to the learners’ needs for vocabulary learning in class, or may not have been designed for specific pedagogical purposes. To bridge this gap, the vocabulary used in some textbook series were analyzed using the vocabulary profiler available in Lextutor.com to identify the lemmas of different frequency levels. The analysis shows that the total number of lemmas in the textbooks was approximately about 58% of the total number of the most frequently used lemmas in the New General Service List (Browne, Culligan, & Phillips, 2013). Further exploration of the lexicogrammatical environment in the textbook corpus discovered lexical items that behave differently with regard to functions and meanings. These findings should provide lexicographers or teachers with useful information about word entries for a relevant learner dictionary to be used in the classroom.

Keywords: lemma; dictionary; vocabulary; concordance

Every teacher and student will agree that having sufficient stock of vocabulary is essential to language learning. It has been widely acknowledged that words and phrases are basic components of language in communication. There are many ways or strategies that teachers can apply to teach words and phrases to students. One of them is to use a dictionary. A large proportion of the vocabulary used in a textbook is usually basic vocabulary. Basic vocabulary is “a list of the most important high-frequency words useful for second language learners of English…” (Browne et al., 2013). This list provides information about the first 1000 most frequently used words (K-1). It also contains the second and third 1000 most frequently used words (K-2 and K-3). This K-1, K-2, and K-3 words should be understood by the students early in English language learning. Schmitt (2000) suggests that at early stages of learning, students should learn about 1000-2000 high-frequency words. A textbook used in schools usually has glosses of vocabulary derived either from the chapters or entire books. While glosses are useful for quick reference of word meanings, the selection may not have been based on frequency decisions but intuitions of the textbook writer.

Learning vocabulary in the classroom generally involves the use of a dictionary as the primary source for definitions and examples of word use that learners meet in the textbooks. It is a common practice for students to use pocket or electronic dictionaries. These dictionaries may or may not be helpful or efficient for classroom learning because such dictionaries do not always offer example sentences that are relevant to the students’ needs. Although this kind of dictionary is convenient, it may not have been designed with appropriate pedagogical purposes, and it may not promote students’ vocabulary knowledge or motivation for learning. Commercial dictionaries are generally used for a general audience. They are not dictionaries designed explicitly for a specific group of students to understand vocabulary in the textbooks being used in schools. A dictionary that students need is the one that they can use to address their real needs for word meaning during English classes, that is, the words that are used in the textbooks.

This study is an attempt to fill this gap to lemmatize the vocabulary in the student textbooks for dictionary entries of basic vocabulary, that is, the most frequently used vocabulary. Such a dictionary can serve as a handy learning tool that is useful and relevant to the students and teachers because it directly addresses students’ learning needs in the classroom; that is to master the material being learned or to pass the tests. This idea may seem to deprive the learners of a broader aim of learning English for communicative purposes, but we can not expect students to develop communicative skills without sufficient vocabulary knowledge.

Despite the teacher qualification in TEFL or TESOL and frequent teacher training to develop teaching skills, English proficiency of our students may not have been satisfactory. This condition appears to be pervasive in EFL contexts such as those in Indonesia (Muhson, 2014), Saudi Arabia (Nezami, 2012), China (Lin, 2002), and Iran.
(Kheirzadeh & Tavakoli, 2012) showing that EFL learners have difficulties in reading comprehension because they lack sufficient vocabulary knowledge. This condition may have been due to the needs of the learners that have not been met in language learning where vocabulary acquisition requires constant exposure of input as frequently described in many studies. Nation (2001), for example, indicates that frequent exposure to new words is necessary for students to understand meanings and learners should be taught how to learn and use new words (Azman, Bhooth, & Ismail, 2013; Bhooth, Azman, & Ismail, 2015).

My observation during teaching practicum shows that the practice of school teachers at the participating schools places a high emphasis on the use of bilingual dictionaries. One strategy for learning vocabulary that is extensively in the literature is the use of bilingual dictionaries. Bilingual dictionaries provide students with examples of sentences with different senses of the vocabulary item. This strategy has recently been used by students to help them understand written or spoken texts.

For vocabulary learning, there are two categories of dictionaries: monolingual dictionary and bilingual dictionary. However, the issue as to which dictionary is the most effective for learning is still debated. Nation (2008), for example, maintains that bilingual dictionaries offer advantages for faster L2 vocabulary learning because the L1 equivalents are provided. A similar conclusion is also mentioned by Folse (2006), Lotus and de Groot (1998) who state that students will have better word retention level if L1 translations are provided. On the other hand, Chan (2004) and White (1997) consider bilingual dictionaries as limited, and they contain rigid or imprecise L1 translations that may not be helpful to learners to develop lexical awareness.

Using the corpus from prescribed textbooks to design a tailor-made bilingual dictionary may provide an answer to vocabulary learning difficulties especially at the beginning stages of learning because students are given an L1 translation of high-frequency words. The studies by Laufer and Hadar (1997), Wu (2005), Marin-Marin (2005) and Amirian and Heshmatifar (2013) show that the students in Taiwan, Mexico, and Iran used and benefited from using bilingual dictionaries. This finding may apply to Indonesian contexts where English is a foreign language. These studies point to the need for a bilingual dictionary that provides L1 equivalents and sufficient exposure of how specific vocabulary items are used in different sentences with different contexts. A bilingual dictionary will provide students with examples of sentences with new words and their translation in L1. Studies by Nation (2001, 2017) and Tsai and Chang (2009) indicate that learners acquire vocabulary much more efficiently using bilingual dictionaries. If students are trained to use an appropriate dictionary correctly, they will be able to read and understand sample sentences that provide illustrations of the shades of meaning of words through multiple exposures, a necessary condition for vocabulary acquisition.

The benefits of using dictionaries have been found positive in enhancing students’ vocabulary knowledge. McAlpine and Myles (2003) clarified that regardless of the type of dictionary, the primary
Purpose of using it is to help learners improve their vocabulary size and increase their awareness of common grammar errors. Incidental vocabulary learning may take place during listening and reading activities with the aid of dictionaries especially helpful to less proficient learners. Luppescu and Day (1993) have studied the potential benefits of improving students' vocabulary knowledge by using bilingual dictionaries appropriately because it can have a substantial positive impact on vocabulary learning and reading development.

Use of concordances for vocabulary learning has been gaining popularity in recent years. Studies on the use of concordances (Al-Mahbashi, Noor, & Amir, 2015; Cobb, 1997; Poole, 2012) have provided information to learners who could benefit from concordance output to enhance their vocabulary knowledge. Concordances can help learners infer meanings and acquire productive vocabulary through multiple usages of vocabulary in authentic contexts. Concordance output can be manipulated to create motivating materials and activities for vocabulary learning that can enhance learners' lexical competence and promote students' autonomy. Some studies on the use of concordancers (Kaur & Hegelheimer, 2005; Poole, 2012; Schmitt, 2000) show that the students learned vocabulary inductively and exhibited improvement in understanding word meanings and were able to transfer their vocabulary knowledge when reading new texts. It appears that learning vocabulary through concordances can lead students to acquire different senses of word meanings and apply this acquired knowledge in reading new texts. Concordances, as summarized by Nation (2001), provide learners with vocabulary in real contexts with rich information not only about word meaning but also a variety of grammatical features that challenge learners to construct generalization and patterns of word usage. The significant role of context in concordance-based vocabulary learning is well recognized (Poole, 2012) since it exposes students to authentic usages of vocabulary in a variety of meaningful contexts. Lin and Huang (2008) state that contexts in concordances offer students with meaning-inferencing activities that are considered to be more efficient than meaning-given activities commonly found in vocabulary look-up activities using pocket dictionaries.

Considering the need for a more relevant dictionary of basic vocabulary and the urgency of acquiring high-frequency words and easy access to a user-friendly online tool (Lextutor.com), this small-scale study attempts to lemmatize the vocabulary items in some English textbooks currently used in Junior High Schools in Indonesia. The textbooks have been recommended by the Indonesian Ministry of Education. The lemmatized vocabulary in this study can be used as a reference to create a dictionary of high-frequency words that are appropriate and relevant to the students at this level with sample sentences selected from the textbooks.

**METHOD**

This study was conducted in a private university in Central Java, Indonesia. It used a documentary method. Documentary methods are the techniques used to categorize, investigate, interpret and identify written documents, whether in the private or public domain such as personal papers, commercial records, or state archives, communications or legislation (Payne & Payne, 2004). In this research, the textbooks were considered as documents, and the vocabulary in the textbooks was taken as data.

The samples of this research were three series of English textbooks for Junior High Schools that have been recommended and documented in the Ministry of Education website (http://bse.kemdikbud.go.id/). The textbooks are presented in Table 1. Note that the vocabulary has been screened for the research.

Each of the sample textbooks was downloaded and converted into word document files. Then, the words were ‘screened’ to remove proper names, numbers, phonetic transcripts, illustrations, and Indonesian words. Thus, only the function and content words necessary for analysis were retained. Finally, the cleaned texts were fed into Lextutor with the following steps.

The first step was data analysis using the Vocabulary Profiler to produce word frequencies (K-1, K-2, K-3) for each textbook. The output shows a list of vocabulary items with different frequencies for each textbook. Since the word frequency lists contain words with their frequency of occurrences (e.g., about [3]), it was necessary to extract the lists to obtain only the lexical items. Thus, this step produced word lists of K-1, K-2, K-3 of each book.

The next step in the analysis was to run Text Compare in the Lextutor to compare the lexical items in K-1, K-2, and K-3 across the books. The result of this comparison was lists of words that were shared as well as unique in the word lists being compared. The shared and the unique word lists were combined and made up all words in K-1, K-2, and K-3 used in all books. However, the combined word lists (K-1, K-2, and K-3) contained lexical items that were not organized. Therefore, the Excel program was utilized to arrange the words alphabetically for the dictionary entries.

**FINDINGS AND DISCUSSIONS**

The steps above produced many pages of word list with 1645 lemmas. A truncated list of the lemmas is presented in Table 2.
Table 1. Textbooks downloaded and their respective vocabulary size

<table>
<thead>
<tr>
<th>Textbook</th>
<th>Publishers</th>
<th>Vocabulary size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bahasa Inggris When English Rings a Bell, Grade VII</td>
<td>Kementrian Pendidikan dan Kebudayaan, 2014</td>
<td>41,354</td>
</tr>
<tr>
<td>Bahasa Inggris When English Rings a Bell, Grade VIII</td>
<td>Kementrian Pendidikan dan Kebudayaan, 2014</td>
<td></td>
</tr>
<tr>
<td>Bahasa Inggris Think Globally Act Locally, Grade IX</td>
<td>Kementrian Pendidikan dan Kebudayaan, 2014</td>
<td></td>
</tr>
<tr>
<td>English in Focus, for Grade VII</td>
<td>Pusat Perbukuan, Departemen Pendidikan Nasional, 2008</td>
<td>65,642</td>
</tr>
<tr>
<td>English in Focus, for Grade VIII</td>
<td>Pusat Perbukuan, Departemen Pendidikan Nasional, 2008</td>
<td></td>
</tr>
<tr>
<td>English in Focus for Grade IX</td>
<td>Pusat Perbukuan, Departemen Pendidikan Nasional, 2008</td>
<td></td>
</tr>
<tr>
<td>Bahasa Inggris, Sekolah Menengah Pertama, Kelas VII</td>
<td>Pusat Perbukuan, Departemen Pendidikan Nasional, 2008</td>
<td>40,203</td>
</tr>
<tr>
<td>Bahasa Inggris, Sekolah Menengah Pertama, Kelas IX</td>
<td>Pusat Perbukuan, Departemen Pendidikan Nasional, 2008</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Truncated word list

<table>
<thead>
<tr>
<th>K-1 word list</th>
<th>K-2 word list</th>
<th>K-3 word list</th>
</tr>
</thead>
<tbody>
<tr>
<td>about</td>
<td>abandon</td>
<td>absorb</td>
</tr>
<tr>
<td>above</td>
<td>able</td>
<td>accent</td>
</tr>
<tr>
<td>accept</td>
<td>abroad</td>
<td>acid</td>
</tr>
<tr>
<td>according</td>
<td>absence</td>
<td>active</td>
</tr>
<tr>
<td>across</td>
<td>absolute</td>
<td>active</td>
</tr>
<tr>
<td>active</td>
<td>accommodate</td>
<td>aeroplane</td>
</tr>
<tr>
<td>actual</td>
<td>accompany</td>
<td>after</td>
</tr>
<tr>
<td>add</td>
<td>achieve</td>
<td>airplane</td>
</tr>
<tr>
<td>address</td>
<td>acknowledge</td>
<td>alien</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Ta’ammeh (2015) states that multiple exposures of new words with their L1 equivalents will facilitate learning because this provision is necessary and efficient (Nation, 2001, 2017; Tsai & Chang, 2009). With a lemmatized word list in place such as the one above, examples of word usages from the textbook corpus can be selected using the concordance facility in the Sketchengine, an online corpus tool available at https://the.sketchengine.co.uk (see Figure 1). The concordance output below illustrates the usages of get in different contexts and the selection of sample sentences can be based on the lexicogrammatical environment of get to provide sufficient exposures for word meanings of get.

Figure 1. The concordance output of get produced by the sketchengine corpus tool.

Here are a few examples sentences with get derived from the lines that can be included to illustrate its meanings in different grammatical contexts.
Astika, Lemmatizing textbook corpus for learner dictionary of basic vocabulary

The lemmatized lexical items in this study were explored to see how a particular word works or functions in relation to other words in its immediate contexts using an online tool called Sketchengine (https://the.sketchengine.co.uk). With this tool, the function of a word or its meaning can be inferred from its lexicogrammatical environment. For example, in this study, the word *have* behaved differently in different environments in the textbook corpus and carries different meanings as indicated in the following sentences.

a. They are tall and *have* dark hair.
   - *have* = to own or possess
b. You always *have* breakfast every morning.
   - *have* = to eat breakfast
c. She will *have* a birthday party on Wednesday.
   - *have* = to arrange a party
d. *We have* to swim to the island.
   - *have* = to swim
   - *must* = to swim
e. *Have* you eaten the food?
   - *have* = a question for a completed act
f. *I have* to take an English course.
   - *have* = to take a course
   - *join* = to take a course

<table>
<thead>
<tr>
<th>Word group</th>
<th>No of lemmas</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-2</td>
<td>708</td>
<td>43</td>
</tr>
<tr>
<td>K-3</td>
<td>279</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>1645</td>
<td>100</td>
</tr>
</tbody>
</table>

Another interesting example in the textbook corpus is the word *take*. It is used in the following phrases with different meanings:

- *take care of* = memelihara. How do you take care of plants?
- *take a bath* = mandi. Take a bath and then have your dinner.
- *take a walk* = jalan-jalan. You are much excited and have decided to take a walk.
- *take a rest* = istirahat. What you need is just take a good rest and drink a lot of fresh water.
- *take turns* = giliran. In your group, you will take turns making a puzzle.
- *take part* = ambil bagian. The government also takes part to increase the function of post offices.
- *take a breath* = ambil nafas. If smoke is around you don't be panic. Take short breaths and crawl.
- *take a medicine* = minum obat. You need to take medicine soon.
- *take a picture* = memotret. The man wants to take a picture of a bird.
- *take actions* = mengambil tindakan. Get ready to...

These examples show that *get* in different grammatical contexts has different meanings in Indonesian and the meanings should be supplied in the entries. In Table 3, the number of lemmas in the sample textbooks have been tabulated in the table below showing the word frequency group, the lemma count of each group and its percentage.

As displayed in Table 2, the number of lemmas (headword) of K-1 is 40% (658 words), K-2 43% (708 words), and K-3, 17% (279 words). In total, the number of lemmas used in the textbooks is 1645, which is approximately about 58% of the total number of K-1, K-2, and K-3 words in the New General Service List (http://www.newgeneralservicelist.org/). This number of lemmas does not include those that fall under the Academic Word Category because this group of words are generally used in academic texts and may not be very useful for students at the junior high school level.

The 1645 lemmas used in the textbooks currently used in the junior high schools in Indonesia should be acquired by the students either through explicit teaching of vocabulary or use of a dictionary. As mentioned earlier, commercially available dictionaries may or may not be convenient or relevant for learning vocabulary in class. Therefore, a tailor-made dictionary based on the list of vocabulary above should be able to meet the students’ needs for learning English using the textbooks mandated by the Indonesian Ministry of Education.

The vocabulary items Table 1 have been organized alphabetically for each frequency level. Since there are three levels (K-1, K-2, and K-3), it is possible that three volumes of dictionaries can be created. It should be noted that the number of words in the K-3 level is not large enough to warrant a sufficiently large size of vocabulary entries. To make up for its small number of entries, it is necessary to include words categorized in the AWL (Academic Word List) group. Although the words in the AWL category are commonly used in academic publications, the students should have at least receptive knowledge of the meanings of those words because they are used in the textbooks. Understanding of these lower frequency words (AWL group) will improve the students’ understanding of the texts in the textbooks. As pointed out by Nation (2001), frequent exposure is needed for students to understand word meanings and use them for productive use of language. Frequent exposure to sample sentences that use a particular word will help students to be aware of differences of word meaning or connotations (Ta’anneh, 2015).

These examples show that *get* in different grammatical contexts has different meanings in Indonesian and the meanings should be supplied in the entries. In Table 3, the number of lemmas in the sample textbooks have been tabulated in the table below showing the word frequency group, the lemma count of each group and its percentage.

As displayed in Table 2, the number of lemmas (headword) of K-1 is 40% (658 words), K-2 43% (708 words), and K-3, 17% (279 words). In total, the number of lemmas used in the textbooks is 1645, which is approximately about 58% of the total number of K-1, K-2, and K-3 words in the New General Service List (http://www.newgeneralservicelist.org/). This number of lemmas does not include those that fall under the Academic Word Category because this group of words are generally used in academic texts and may not be very useful for students at the junior high school level.

The 1645 lemmas used in the textbooks currently used in the junior high schools in Indonesia should be acquired by the students either through explicit teaching of vocabulary or use of a dictionary. As mentioned earlier, commercially available dictionaries may or may not be convenient or relevant for learning vocabulary in class. Therefore, a tailor-made dictionary based on the list of vocabulary above should be able to meet the students’ needs for learning English using the textbooks mandated by the Indonesian Ministry of Education.

The vocabulary items Table 1 have been organized alphabetically for each frequency level. Since there are three levels (K-1, K-2, and K-3), it is possible that three volumes of dictionaries can be created. It should be noted that the number of words in the K-3 level is not large enough to warrant a sufficiently large size of vocabulary entries. To make up for its small number of entries, it is necessary to include words categorized in the AWL (Academic Word List) group. Although the words in the AWL category are commonly used in academic publications, the students should have at least receptive knowledge of the meanings of those words because they are used in the textbooks. Understanding of these lower frequency words (AWL group) will improve the students’ understanding of the texts in the textbooks. As pointed out by Nation (2001), frequent exposure is needed for students to understand word meanings and use them for productive use of language. Frequent exposure to sample sentences that use a particular word will help students to be aware of differences of word meaning or connotations (Ta’anneh, 2015).

The lemmatized lexical items in this study were explored to see how a particular word works or functions in relation to other words in its immediate contexts using an online tool called Sketchengine (https://the.sketchengine.co.uk). With this tool, the function of a word or its meaning can be inferred from its lexicogrammatical environment. For example, in this study, the word *have* behaved differently in different environments in the textbook corpus and carries different meanings as indicated in the following sentences.

a. They are tall and *have* dark hair.
   - *have* = to own or possess
b. You always *have* breakfast every morning.
   - *have* = to eat breakfast
c. She will *have* a birthday party on Wednesday.
   - *have* = to arrange a party
d. *We have* to swim to the island.
   - *have* = to swim
   - *must* = to swim
e. *Have* you eaten the food?
   - *have* = a question for a completed act
f. *I have* to take an English course.
   - *have* = to take a course
   - *join* = to take a course

<table>
<thead>
<tr>
<th>Word group</th>
<th>No of lemmas</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-2</td>
<td>708</td>
<td>43</td>
</tr>
<tr>
<td>K-3</td>
<td>279</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>1645</td>
<td>100</td>
</tr>
</tbody>
</table>

Another interesting example in the textbook corpus is the word *take*. It is used in the following phrases with different meanings:

- *take care of* = memelihara. How do you take care of plants?
- *take a bath* = mandi. Take a bath and then have your dinner.
- *take a walk* = jalan-jalan. You are much excited and have decided to take a walk.
- *take a rest* = istirahat. What you need is just take a good rest and drink a lot of fresh water.
- *take turns* = giliran. In your group, you will take turns making a puzzle.
- *take part* = ambil bagian. The government also takes part to increase the function of post offices.
- *take a breath* = ambil nafas. If smoke is around you don't be panic. Take short breaths and crawl.
- *take a medicine* = minum obat. You need to take medicine soon.
- *take a picture* = memotret. The man wants to take a picture of a bird.
- *take actions* = mengambil tindakan. Get ready to...
As described by Anderson and Corbett (2009, p. 63), the language used in the textbooks may not reflect authenticity in the real sense, but it may resemble real use of English.

Further investigation of the textbook corpus found that sense of word meaning may differ ‘according to the mode of production, that is, whether a text is spoken or written’ (Anderson & Corbett, 2009, p. 63), or the part of speech that follows. Here is a selection of four concordance lines retrieved from the textbook corpus. Note that the language used in the textbooks may not reflect authenticity in the real sense, but it may resemble real use of English.

**Example Sentences:**
- *really help* to plan and remind us.
- *I love you*.
- *I love you*.
- *I love you*.

These are useful expressions as dictionary entries, and learners can learn the expressions as small chunks of English especially for speaking skill development.

Further exploration on the textbook corpus for this study shows that there are many fixed expressions that can be used as dictionary entries. The following examples are a few among many fixed expressions used in the textbook corpus along with their frequency of occurrences from high to low:

- **Excuse me** (63)
- **Good morning** (60)
- **How are you?** (45)
- **Nice to meet you** (36)
- **Of course** (34)
- **Would you like to ...** (23)
- **I am sorry** (23)
- **Here you are** (21)
- **I am sorry to hear that** (20)

CONCLUSION

This study explored a corpus of English textbooks for use in Indonesian junior high schools to provide information for dictionary entries of high-frequency words. The analysis indicates that the lemmatized word list covers as much as 58% of the high-frequency vocabulary in the New General Service List. The inclusion of the lemmas in this study into a dictionary for the students at this level could be useful for classroom learning using the textbooks because the dictionary entries and sample sentences can be selected from the textbooks being used. However, this amount of word coverage may not be sufficient to equip learners with the necessary knowledge of high-frequency words at the junior high school level.

In addition to the needs to provide meanings of the individual word in the L1, the corpus analysis in this study has also pointed out the needs to include specific words such as *have* and *take* that may carry different meanings when such words collocate with other words. As described by Anderson and Corbett...
knowledge of collocation can help learners infer subtle differences of meanings and usages which may not be evident from meanings of individual words that make up the collocations. Knowledge of set expressions or pragmatic routines should also be useful for dictionary entries. By using the Sketchengine corpus tool, this study identified some routines such as Nice to meet you, or Would you like to... in the corpus that students need to learn and use to develop their sociopragmatic competence. A lot more examples of such routines can be explored in the corpus for dictionary entries. With easy access to online software programs for language learning, English educators will have the necessary tools to elevate teaching and learning burden and make language learning in class more enjoyable. It is worth noting that this study did not cover all English textbooks recommended by the Ministry of Education. The corpus would have been larger if additional textbooks had been included in the samples. Although the textbooks were sampled from those currently being used in Indonesia, a similar study could be carried out with other textbooks used in other EFL settings to provide more information for dictionary entries of basic vocabulary.

REFERENCES


McAlpine, J., & Myles, J. (2003). Capturing phraseology in an online dictionary for advanced users of English as a second


Sketchengine. (n.d.). Available at https://the.sketchengine.co.uk


