

A CORPUS LINGUISTICS INVESTIGATION OF TWO NEAR-SYNONYMOUS

WORDS: *RICH* AND *WEALTHY*

Dwi Yulianto Nugroho

Pelita Harapan University

dwi.nugroho@uph.edu

Abstract

Many language learners, even language teachers, consider two words are synonymous. In fact, some words are not really synonymous. This sometimes triggers questions among language learners. They are confused whether two considered-synonymous words can be used interchangeably in certain contexts. Their confusion will appear more when they should describe something. Therefore, this study tries to find out phenomena around two words which are considered synonymous, namely rich and wealthy. Using British National Corpus (BNC) as the data which were analysed using the software Sketch Engine (SkE), this corpus linguistics based study found out that rich and wealthy are near-synonyms because they have identical sense but they are not substitutable for certain contexts. In addition, rich is more commonly used by people to describe both human and non-human beings. The results of the study imply that corpus linguistics provide benefits for English language teaching as we can really analyse language using authentic facts compiled in a corpus. A language teacher can also use corpus linguistics to teach vocabulary. In addition, a sociolinguistics phenomenon found out in this study implies that investigating two considered-synonymous adjectives can help language learners understand how people use the adjectives in socialising with other people.

Keywords: *Corpus Linguistics, near synonyms, British National Corpus, Sketch Engine*

INTRODUCTION

When it comes to learning about descriptive text, using adjective becomes a dilemmatic issue for language learners. If they cannot choose a correct adjective, their description might not be accurate. Thus, they must be able to select appropriate adjectives to describe something. As the results, most of my students will usually ask me some question about whether two or three words are synonymous and interchangeable in the same context.

One day, a particularly critical student asked me if *rich* and *wealthy* are synonymous. *Rich* and *wealthy* are quite common to be used when we want to

describe a person and of course the student wanted to use those words correctly in a sentence. Therefore, as a teacher, I had to give the correct answer. Answering such a question based on my intuition might be correct, but it is tricky. When I was not sure with my intuition, looking up the words in a dictionary to find a more convincing answer could be the option. According to the Merriam-Webster Online Dictionary and Thesaurus, *rich* and *wealthy* are both adjectives. The dictionary tells us that *rich* is a condition of 'having abundant possessions and especially material wealth', whereas *wealthy* means 'having wealth'. From the definition from the dictionary, it seems that *rich* and *wealthy* are synonymous because both are talking about wealth. In addition, they can be used interchangeably as in the following sentences:

1. Arland is a **rich** person.
2. Arland is a **wealthy** person.

Both sentences indicate that **Arland has wealth**. Moreover, the Indonesian language translation of those words will be *Arland adalah orang kaya*. However, are *rich* and *wealthy* really synonymous? Could we use those words to describe a non-living thing, such as fruits? Could we say *this fruit is wealthy in vitamins*? Those questions surely trigger language teachers and learners to find accurate data of whether *rich* and *wealthy* are synonymous and interchangeable. This indicates that deciding two words synonymous is simple. Lyons (1995) mentions two types of synonyms, namely absolute-synonyms and near-synonyms. To be called absolute-synonyms, two or more expressions must be identical, have same meanings in all contexts, and if we view the expressions from semantics aspect, they must be equivalent on all meaning dimension. From these, it is quite difficult for most of synonyms to be called absolute synonyms, and likely to be near-synonyms. Inkpen and Hirts (2006:223) add that near-synonyms are words which are 'not fully intersubstitutable, but rather than vary in their shades of denotation or connotation, or in the components of meaning they emphasise; they may also vary in grammatical or collocational constraints'.

Therefore, based on those theories, *rich* and *wealthy* are near-synonyms as they are not fully intersubstitutable. As the results, we need to be able to use them in appropriate context. To do that, we need to analyse the words' discourse prosody,

which are also known as semantic prosody or semantic preference. According to Stubbs (2001:65), semantic preference is ‘the relation, not between individual words, but between a lemma or word-form and a set of semantically related words’. Baker (2006:87) argues that ‘semantic preference denotes aspects of meaning which are independent of speaker, whereas discourse prosody focuses on the relationship of a word to speakers and hearers and is more concerned with attitudes’. Partington (2004) adds that semantic preference is more related to affective factors and collocation is closely tied with the phenomenon. Therefore, in the study of near-synonyms, we need to consider collocation. According to Bauer (2003:326), a collocation is ‘a habitual co-occurrence of two or more items, which may be syntactic (as in *dry wine* or *turn left*) or morphological (as in *bishop-ric road works*)’. Furthermore, McEnery and Wilson (2001) add that we can analyse the sense of a word through collocation to help us determine broad meaning similarity as well as whether two or more expressions are synonymous. From these, we can see that analysing two or more synonymous words is not easy and depending on our intuition and dictionary meaning is not enough. In the other words, we need to find reliable data to support our analysis.

To get reliable evidence, we can take advantage from the advent of corpus linguistics. Although Hunston (2002) points out that corpus linguistics has a drawback as it will never be truly representative of a language, some experts believe that using corpus linguistics in language study is still beneficial. Finegan (2004:574) argues that corpus is ‘a representative collection of texts, usually in machine-readable form and including information about the situation in which each text originated, such as the speaker or author, addresses, or audience’. It means that corpus represent the language that we use in real context of communication. Furthermore, according to McEnery et al (2006), corpus linguistics is a methodology that can be used by lexicographer to investigate language, including near-synonymous words, such as *rich* and *wealthy*. McEnery and Wilson (2001) stated that corpus linguistics provide empirical data which: ‘enable the linguist to make statements which are objective and based on language as it really is rather than statements which are subjective and based upon the individual’s own internalised cognitive perception of the language’ (McEnery and

Wilson, 2001:87). In addition, Biber et al. (2002:26) add that ‘one of the advantages of corpus-based research is that the corpus can be used to show all the contexts in which a word occurs. From these contexts, it is often possible to identify the different meanings associated with a word’. Therefore, we can conclude that analysing two near-synonymous words *rich* and *wealthy* using corpus linguistics will end with reliable data.

Some researchers have investigated near-synonyms using corpus linguistics. Yang (2016) conducted a corpus-based comparative study of *learn* and *acquire*. Furthermore, based on the Longman-Lancaster Corpus, Biber et al (2002) analysed seemingly synonymous words *big*, *large*, and *great* from two registers, namely fiction texts and academic prose. The study concluded that ‘corpus-based analysis can be used to show how each adjective has its own preferred collocates, different preferred senses, and different distribution across registers’ (Biber et al, 2002:51). Liu’s study (2010) revealed the benefits of using corpus linguistics for language study. By using the Corpus of Contemporary American English, Liu investigated five near-synonyms, namely *chief*, *main*, *major*, *primary*, and *principal*. Moreover, a corpus-based study conducted by Chung (2011) revealed useful data in which *create* and *produce* often share two senses which are overlapping although they are usually used as a bare infinitive and in the *-ed* forms. Furthermore, a study on English synonyms conducted by Phoocharoensil (2010) showed that corpus-based analysis contributes more to language study compared to an investigation which was based on learners’ dictionaries. These studies indicate that investigating near-synonyms using corpus-linguistics is surely beneficial. As an attempt to use corpus linguistics for a language study, this paper describes whether *rich* and *wealthy* are synonymous, and if yes, to what extent they are synonymous. It is expected that this paper can give information about how they are used in a sentence appropriately.

METHODOLOGY

British National Corpus (BNC) was chosen as the source of data which was then analysed using the Sketch Engine. According to Kilgarriff et al. (2014:7), Sketch Engine is ‘a leading corpus tool, widely used in lexicography’. Then, a tool in the

Sketch Engine named the Word Sketch was chosen to examine the corpus related to *rich* and *wealthy*. Through the Word Sketch tool in Sketch Engine, we can get a one-page summary of not only word's grammatical but also their collocational behaviour, as shown in Figure 1.

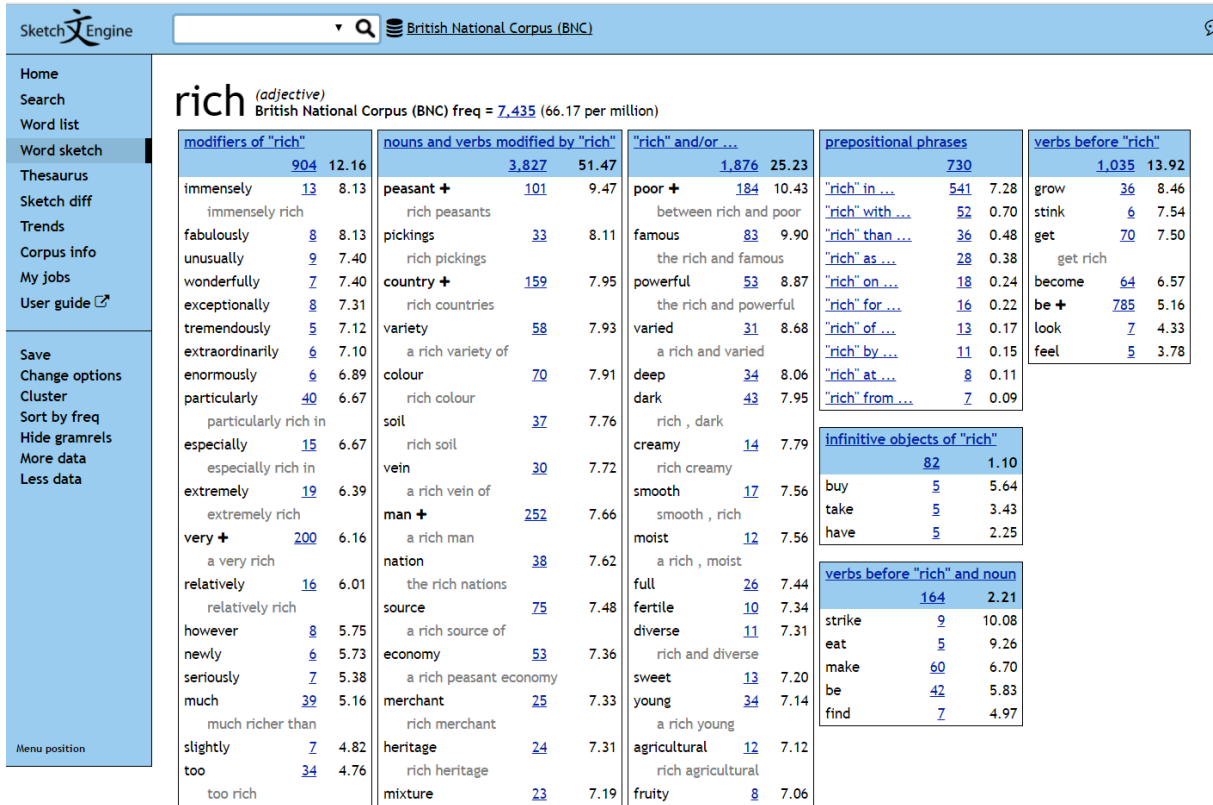


Figure 1. Word Sketch for English *rich*, adjective (from BNC)

Figure 1 provides a one-page summary of collocation behaviour of the word *rich*. We can see that *rich* can be used to modify nouns. In order to analyse the difference between *rich* and *wealthy*, I focused on the nouns which are modified by *rich* and *wealthy* because of of them are adjectives. Liu (2010) stated that the main function of adjective is to modify nouns. In order to get more interesting findings such as the phenomenon of using double adjectives in a sentence, I will also compare *rich* and *wealthy* with the other adjectives in the corpus. I will then choose ten most common nouns or adjectives which go with *rich* and *wealthy* together in a sentence. To select those ten most commons nouns or adjectives, *logDice* score. According to Rychly (2008:9), 'The *logDice* score has a reasonable interpretation, scales well, on a different corpus size, is stable of subcorpora, and the values are in reasonable range'. Thus, the results will be more reliable.

RESULTS

It is necessary to figure out the frequency of two near-synonymous words in a corpus. From the frequency, we can see how many times two words are used in communication. The following table shows the frequency of *rich* and *wealthy* in BNC.

Table 1. The frequency of *rich* and *wealthy* in BNC

Word	Frequency
Rich	7,435 (66.17% per million words)
Wealthy	1,537 (13.68% per million words)

From Table 1, we can see that *rich* is more commonly used in both spoken and written communication than *wealthy*. However, knowing this fact does not make us understand the meaning and specific usage patterns of the adjectives (how they are correlated with the other nouns and adjectives). Therefore, we need to see most common nouns and adjectives which collocate with *rich* and *wealthy*. Table 2 below shows us the ten most common nouns which are modified by *rich* and *wealthy*.

Table 2. Ten most common nouns modified by *rich* and *wealthy*.

No	Rich	Wealthy
1	Peasant	Merchants
2	Pickings	Industrialist
3	Country	Landowner
4	Variety	Widow
5	Colour	Businessman
6	Soil	Individual
7	Vein	Patron
8	Man	Benefactor
9	Nation	Peasant
10	Source	Aristocrat

From Table 2, we can see that *rich* has broad usage than *wealthy*. While *wealthy* are commonly used to describe persons such as merchants and industrialist, *rich*

can be used to describe persons and any other things such as country and colour. From these findings, we can now understand why *rich* has greater number compared to *wealthy*. People use *rich* in more various description of something. From Table 2, we can also figure out that people commonly describe *peasant* using both *rich* and *wealthy*. From this case, it seems that that *rich* and *wealthy* are absolute-synonyms because they share the same meaning in the same context. However, if we take a look at The Merriam-Webster Online Dictionary, *peasant* is defined as ‘a member of a European class of persons tilling the soil as small landowners or as labourers’. This indicates that *peasant* is commonly known in European countries only as this is contextual. Thus, this implies that teachers from non-European countries should not take *rich peasant* or *wealthy peasant* as an example of collocation when they are teaching about describing people in the classroom.

In the next discussion, we will see how *rich* and *wealthy* are used with another adjectives. Table 3 below presents ten adjectives which mostly and commonly go with *rich* and *wealthy*.

Table 3. Ten most common adjectives which go with *rich* and *wealthy*.

No	Rich	Wealthy
1	Poor	Influential
2	Famous	Privileged
3	Powerful	Powerful
4	Varied	Respectable
5	Deep	Landed
6	Dark	Middle-class
7	Creamy	Wise
8	Smooth	Able
9	Moist	Handsome
10	Full	Successful

From Table 3, we can see that *rich*, although it is not always, is commonly contrasted with adjectives which have negative sense, such as *poor* and *dark*. Meanwhile, we do not see this phenomenon when we use *wealthy* to go with

another adjective. This indicates that when we want to contrast someone who has wealth or positive situation with any situation which brings unhappiness, we may use *rich* instead of *wealthy*.

From Table 3, when *rich* goes with other adjectives, we can see that people use *rich* to describe non-human beings as shown by the following figures taken from BNC.

Written bo... Holland. </p><p> Restaurants: Dutch cuisine is **rich** and **creamy** and comes in large portions;
Written bo... professor commented, The dessert is smooth, **rich** and **creamy** in texture - qualities we subconsciously
Written bo... Patrick Campbell. The voice needs to be **rich** and **varied** , flirtatious and worldly wise
Written bo... by those historians who have surveyed the **rich** and **varied** political ideologies of the

Figure 1. *Rich* goes with *creamy* and *varied*

Figure 1 shows that *rich* modifies *cuisine*, *dessert*, *voice*, and *political ideologies*. This confirms the findings that when people want to describe non-human beings, they prefer to use *rich* than *wealthy*. Therefore, teachers can explain it to their students that *rich* has broader usage compared to *wealthy*.

From Table 3, we can also investigate interesting facts related to sociolinguistic phenomena. As what McEnery and Wilson (2001) have asserted, corpus linguistics can be used by sociolinguists to analyse any social phenomena in the society. From Table 3, we can see that when they are used to describe humans, both *rich* and *wealthy* usually go together with some words which show power, for instance *influential* and *powerful*. Figure 2 below shows the extracts from the corpus.

Written bo... , because he is poor and his opponent is **rich** and **powerful** , because juries are corrupt
Written bo... increasingly becoming the servants of the **rich** and **powerful** , but he did not believe that
Written bo... William married Isabella Grant, daughter of a **wealthy** and **influential** family. His father-in-law
Written bo... He enjoyed the company of **influential** and **wealthy** people, including royalty. He had a fine

Figure 2. *Rich* and *wealthy* to modify human beings

Figure 2 reveals social phenomenon in our society. From that, we can see that people who are rich or wealthy are those who are also powerful and influential. From Table 3, we can also get *wealthy* and *successful* and also *wealthy* and *privilege*. This indicates that rich or wealthy people are more respected in the society as they are considered successful and powerful. Thus, they are more influential as the voice from the poor is not really powerful. This can trigger social studies on how people nowadays see rich and poor people in the society and how rich people treat the poor.

CONCLUSION

From the data and analysis, we can conclude that *rich* and *wealthy* are near-synonymous words because they have identical senses in respect of central semantic traits but they are not substitutable in certain contexts. Furthermore, *rich* has broader usage than *rich*. Therefore, if we want to describe non-human beings, *rich* is more preferred as *wealthy* is more commonly used to describe human beings only.

Corpus linguistics is beneficial for many people, such as language teachers and learners, lexicographers, and sociolinguists. When language teachers want to ask their students to use corpora in language teaching, teachers should be aware that the corpus might be contextual, as the shown by the case of *wealthy peasant*. However, we can see some potentials of using corpus linguistics in vocabulary teaching as they provide authentic usage of language. Furthermore, using corpus linguistics to analyse the use of adjective words in the real communication can help people see what is happening in our society. Therefore, corpus linguistics provides advantages to sociolinguists to investigate social phenomena.

REFERENCES

- Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Bauer, L. (2003). *Introducing linguistics morphology*. Washington DC: Georgetown University Press.
- Biber, D., Conrad, S., and Reppen, R. (2002). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Chung, S-F. (2011). A corpus-based analysis of 'create' and 'produce'. *Chang Gung Journal of Humanities and Social Sciences*. **4**(2), pp. 399-425.
- Finegan, E. (2004). *Language: its structure and use*. Boston: Wadsworth, Cengage Learning.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Inkpen, D., and Hirst, G. (2006). Building and using a lexical knowledge base of near synonym differences. *Computational Linguistics*. **5**(3), pp. 223-262.
- Kilgarriff, A. et al. (2014). The Sketch Engine: ten years on. *Lexicography ASIALEX*. **1**(1), pp. 7-36.

- Liu, D. (2010). Is it a chief, main, major, primary, or principal concern? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics*. **15**(1), pp.56-87.
- Lyons, J. (1995). *Linguistics semantics*. Hove and London: Lawrence Erlbaum Associates, Publishers.
- McEnery, T., and Wilson, A. (2001). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T. et al. (2006). *Corpus-based language studies: an advance resource book*. London: Routledge.
- Merriam-Webster Online Dictionary (Available from: <https://www.merriam-webster.com/dictionary/rich>, <https://www.merriam-webster.com/dictionary/wealthy>) [Online]. [Accessed on Thursday, May 11th, 2017].
- Partington, A. (2004). Utterly content in each other's company: semantic prosody and semantic preference. *International journal of corpus linguistics*. **9**(1), pp.131-156.
- Phoocharoensil, S. 2010. A Corpus-Based Study of English Synonyms. *International Journal of Arts and Sciences*. **3**(10), pp.227-245.
- Rychly, P. (2008). A lexicographer-friendly association score. In: Sojka, P. & Horak, A. (Eds). *Proceeding of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008*. [Online]. Brno: Masaryk University, pp. 6-9. [Accessed on Wednesday, June 7th, 2017]. Available from: <http://www.fi.muni.cz/usr/sojka/download/raslan2008/13.pdf>,
- Sketch Engine. (2017). Statistics used in Sketch Engine. [Online]. [Accessed on Wednesday, June 7th, 2017] Available from: <https://www.sketchengine.co.uk/documentation/statistics-used-in-sketch-engine/>
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Yang, B. (2016). A corpus-based comparative study of *learn* and *acquire*. *English Language Teaching*. **9**(1), pp.209-220.
<http://dx.doi.org/10.5539/elt.v9n1p209>