THE LEXICOGRAMMAR OF <ADJ + TEA> CONSTRUCTION IN THE CORPUS OF CONTEMPORARY AMERICAN ENGLISH: HOW DO YOU LIKE YOUR TEA?

Prihantoro

Universitas Diponegoro

[prihantoro2001@yahoo.com](mailto:prihantoro2001@yahoo.com)

**Abstract**

One of the functions of adjectives is attributive to its head. This study is aimed at describing the adjectives that frequently co-occur with TEA. Tea is one of the beverages that are familiar with people all around the world. Adjectives here take a crucial role in describing the TEA. I here retrieved the top 100 adjectives that collocate with TEA in the corpus of Contemporary American English (COCA), and mapped their semantic and grammar class. The results are thematically divided into three parts. In terms of frequency, the phrase <iced tea> is the most frequent pattern, which took 860 hits. The last (100th) pattern is <giant tea>, which only got 4 hits. In terms of grammar, the retrieval indicates that adjectives in <ADJ + TEA> construction may take the following forms: stem, derivates, hyphenated and spaceless compound. And the last, in terms of semantic categorization, +MANIPULATION, +TASTE, and +ORIGIN are some of the most frequent classes, which are shown by the following examples respectively; <iced tea>, <mint tea>, <EnglishTea>. This authentic data driven result might assist learners of English as a Foreign Language as adjectives that co-occur with tea in their native languages might be different to some extent.

Keywords: COCA, tea, adjective, attributive, lexicogrammar

1. Introduction

One of the widespread commodities around the world is tea. Tea is originally from Asia (Hasimoto, 2001), but consumed by people all around the world. Although it seems to be a universal beverage, tea processing and the way people enjoy tea these days may vary all around the world. Language as a cultural artifact takes crucial importance here to communicate how people process and taste tea.

This research is centralized on lemma TEA. The focus of this research is to reveal how TEA is described by English speakers (spoken or written). Readers will be presented by common and frequently used expressions on how English speakers describe TEA.

I here look into how TEA is attributively described in the Corpus of Contemporary American English (COCA) by means of adjectives. In the retrieval, I employed <ADJ + TEA> construction and the top 100 results were analyzed. Morphological and semantic categorizations will be presented in the discussion section in this paper.

1. Previous Studies

There are other languages, where the equivalent of ‘tea’ is similar; *teh* (Hebrew, Indonesian, Malay), thee (Dutch), te (Danish) and etc. Another form, very different but also widespread is ‘chai’; *shai* (Arabic), *cha* (Chinese, Korean), *chai* (Azerbaijani) and etc. See the distribution in figure 1:

Figure 1. The Distribution of *tea* and *chai* Similar Forms

(http://www.languagesoftheworld.info)

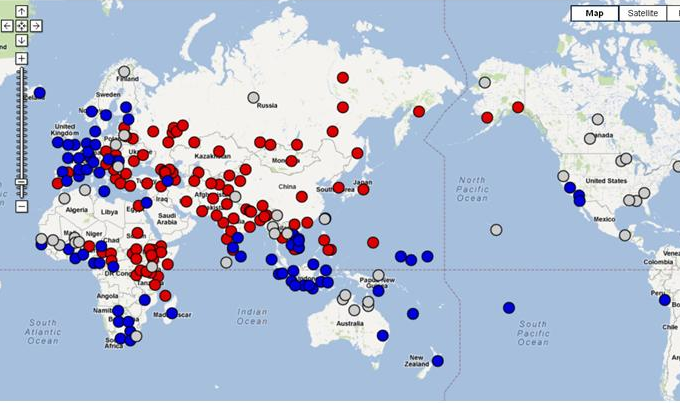


Figure 1 illustrates the distribution of similar forms of *tea* and *chai* around the world. The blue (darker dot) is similar to *tea* and the red ones (brighter) are similar forms to *chai*. The survey, however, do not focus on analyzing the word specifically in each language. However, it is useful to understand how the word was adopted from one language to another.

Sarmah (2015) in his paper about language maintenance shows how the presence of tea pickers/gardeners is influential in maintaining one of the vernacular languages in India. The decreasing number of tea pickers positively correlates to the worsening vitality of the language maintenance.

Ludeling & Kyto (2009) illustates ‘strong tea’ as a frequent phrases from corpus processing perspectives. They believe that the more frequent the collocates, the higher MI (Mutual Information) score will be. MI is on of the scores used in statistical machine translation to provide automatic translation and other natural language processing functions.

The closest to this study in terms of object and method is the work of Laybutt (2009). He compared two expressions ‘a (cup|mug) of tea’. By using corpus approach to 13 corpora and several representative english dictionaries, he argued that the two expressions are different in terms of semantic and syntactic. The use of ‘cup’ seems to be more natural and created by culture/habit, while ‘mug’ is more on the grammaticalization and syntactic force.

Unlike Laybutt, in this study I focus on COCA (American English). It must be understood that either the driving force is culture or syntax, they are under the frame of American English speakers. Different from Laybutt who focused on the noun (specifically containier), my study is focused on adjectives that is used attributively to describe TEA. How Laybutt’s and my study can complement one and each other will be discussed in the discussion.

1. Methodology

This research focuses on COCA (Davies, 2008) data for some reasons. As it is well understood, in the world of corpus linguistics, data representativeness is an important issue. As for COCA, it is claimed to be one of the most representative, well balance over text types and largest corpora so that reliable to be a monitor corpus. Second, corpus data in COCA is already annotated allowing POS based retrieval. Three, COCA interface allows the results to be presented by means of hits (number) or visualization (graphics, chart etc). And the last one is its access. COCA is freely accessible as long as you have internet connection. You do not have to pay.

In the retrieval box, I wrote <ADJ + TEA> regular expression, which correspond to COCA interface syntax. The interface then computed the data, retrieving all <ADJ + TEA> strings. It retrieved the orthographic form of TEA (which happens to be the same ‘tea’) preceded by any token marked as adjective (ex: sweet/A, bitter/A, nice/A). Each adjectives is presented by its frequency. I extracted the top 100 (100 most frequent) adjectives that collocate with tea. These adjectives were then categorized morphologically and semantically. They are discussed in the next section.

1. Discussion

4.1 Morphology

While the basic function of adjectives here is clearly for attribution (Azar, 2002), the morphology vary. A word might come as a result of inflection, derivation, compounding, or other word formation process (Aranoff & Fudeman, 2011). Simple adjectives are highest in frequency in the result of the retrieval (57). Some of the examples are <hot, cold, sweet, black>. Other adjectives, which are also significant in frequency is derived adjectives, which originated from other word class, and derived by suffixation such as verb <poured-pour> and noun <traditional-tradition>. Some of the affixed adjectives are complex such as <decaffeinated, unsweetened>. Note that although suffix –ed is attached, but the function is not inflectional but derivational. Also not all suffixes change the word class (de-, un-). They are, however, derivational in function .

By using <ADJ + TEA> construction, I did not expect to obtain any compound. However, six compounds were included in the result. If we understand the way COCA interface works, it actually makes sense. The syntax I wrote retrieved two token. In English, token is defined as character strings separated by space (simply say). Therefore, hyphenated and spaceless compound adjectives like *anti-tax* or *lukewarm* were also included in the retrieval. As a human, we understand that although each

token is simple in form, but they are actually multi-word units (Paumier, 2014) or compounds. The machine does not understand this and go straight with what is programmed. Table 1 summarize the adjectives forms:

Table 1. Word Formation Statistic of top 100 Adjectives in <Adj + Tea> Construction

|  |  |  |
| --- | --- | --- |
| Adjective Stems | | 57 |
| Derivational Adjectives | | 37 |
| Compound Adjectives | Spaceless | 1 |
| Hyphenated | 5 |

4.2 Semantic

The meaning of the construction is divided into two: compositional and uncompositional. Sinclair’s *Reading Concordance* (2003) has shown the reading of multiword units corpus database into two. One is where the meaning is predictable via each word. And another is when we have to consider the words as one unit, where meaning is unpredictable via its constituents.

Compositional Meaning

Compositional here is defined in terms of the shift. For instance, sweet tea is compositional where it refers to a cup of tea that is sweet in taste. Therefore, the meaning is deductible from the each lexical item, which composes the pattern. For instance, Sweet, in ‘sweet tea’, does not refer to attitude as in ‘how sweet of you to pay a visit’. It can be traced compositionally as flavor.

Table 2. Compositional Meaning of <Adj + Tea> Construction

|  |  |
| --- | --- |
| Hot | tea |
| Sweet |
| Mint |
| National |
| Chinese |
| decaffeinated |

Table 2 shows the some common modifiers distribution for <tea> in COCA. Although the meaning is deductible from the composition, relation of head and modifier may vary. <hot tea> is <tea> THAT IS <hot>. This may apply for <sweet> although the first one refers to temperature and another one refers to flavor. As for <mint> the relation is <tea> MADE OF <mint> indicating also flavor. The flavor however is specified to resemble mint. <national tea> specifies national identity, while <Chinese> is the ORIGIN. <decaffeinated> refers to PROCESS.

However, delineating the meaning shift is not easy. Consider <green tea> for example. Tea is normally brown in color, as it has undergone several processes. They are green only when they are still in plant form. A cup of green tea is also not green. The spectrum would still categorize it as brown, but a little lighter than common tea. Another example is herbal tea. We are aware that tea is also one of the herbs. The naming of herbal tea is based on the fact that other ingredient is also added.

On above cases, meanings shift is not absent. They are present, but very soft, and people can still understand the literal meaning of the constituents that compose the construction. Therefore, the construction is still categorized as compositional.

Uncompositional Meaning

A <white house> can mean a house that is white, or the office of US president. Meaning is compositional in the first, but uncompositional by the later. The meaning of ‘white house’ as the office of US president is not constructed by the relation of each lexical meaning, but by social convention first, then acknowledged later by language science under the name of compound.

Some of the expressions are not compositional; but they are not many. Take a look at *mad tea*  and *pretend tea*. The phrase *mad tea* actually refers to one of the facilities in the amusement park, where children ride big swinging teacup. *Pretend tea* is a part of *pretend tea party* where kids are pretending to be a tea maker.

Figure 2. Mad Tea Party in an Amusement Park

[](https://en.wikipedia.org/wiki/File:TeacupsMadTeaParty_wb.jpg)

Most Frequent Adjectives

Table 3. Top Ten Adjectives in <Adj + Tea> Construction

|  |  |  |  |
| --- | --- | --- | --- |
| ADJECTIVES | NODE | FREQ | SEMANTIC |
| Iced | Tea | 860 | +Temperature |
| Green | Tea | 708 | +Color |
| Hot | Tea | 249 | +Temperature |
| Herbal | Tea | 190 | +Manipulation |
| Black | Tea | 174 | +Color |
| Sweet | Tea | 121 | +Taste |
| High | Tea | 95 | +Measurement |
| Mint | Tea | 78 | +Taste |
| Japanese | Tea | 56 | +Origin |
| Hybrid | Tea | 39 | +Manipulation |

Table 3 shows the adjectives that collocate with tea. First column serves top ten adjectives that collocate with the node, <tea>. Frequency is presented on the third column. Semantic classification of the adjectives are served on the last column

The descending hits shown by the table is an indication that <iced tea> and <green tea> have shown significant statistic; 860 and 708 respectively. Between 100-250, we have <hot, herbal, black, sweet + tea>. Other lexical items such as <sweet, high, mint, Japanese, hybrid> are lower than 100 hits.

Among the top ten adjectives, none of them is attributing quality in general like ‘good’, ‘nice’ or ‘bad’; which means that not all adjectives may collocate with <TEA>. They are specific to special adjective classes such as +TEMPERATURE (*iced, hot*), +COLOR (*green*, black), +MANIPULATION (herbal, hybrid), +TASTE (mint, sweet), +ORIGIN (Japanese), +MEASURE (high).

While the lexical aspect suggests that certain semantic classes adjectives tend to collocate with <TEA>, the morphology of the lexical items is also unique. Besides, canonic adjective stems (green, sweet), some of the adjectives are derived (iced, herbal). However, compounding is absent in the top ten adjectives. The closest one is *lukewarm tea, which* is on 57th place.

Of the top ten <adj+tea>, most of the meaning are compositional; the meaning is deductible by the constituents. However, there is a construction where the meaning is not compositional, which is *high tea*. High tea might refer to refreshments (Brooks, 2014) served in the afternoon, or sometimes called afternoon tea.

1. Conclusion

Adjective is a very strong modifier for a noun, as it may describe the quality of the noun, as its head. As the node of the construction studied in this research is <tea>, the results are <Adj + tea> construction. Adjectives that collocate with <tea> might be categorized as simple adjectives (the canonic/lemmatic form), inflected adjectives, and derived adjectives (including ones that are derived from another word class). Queries designed to obtain <adj+tea> construction in this study retrieved also compound adjectives. This came as a result that compounding might be performed without space, or with hypen {-}. COCA tokenization module, as common as another English module, use space to mark token. Although they are compound, the are not separated by space, therefore analyzed as a token. But in fact, English speakers understand that they are multiword unit.

The adjectives are also categorized to different semantic classes. In terms of meaning, almost all of the construction are compositional; the meaning is deductible from the constituents. There are some fixed expressions where the semantic is not compositional. However, these expressions are not high in terms of frequency, except for *high tea*. *High tea* , which is afternoon refreshment, rank 7, which is in top ten most frequent adjectives.

# References

Aranoff, M., & Fudeman, K. 2011. *What is Morphology.* Cambridge: Cambridge University Press.

Azar, B. S. 2002. *Understanding and Using English Grammar 3rd Edition.* New York: Longman.

Brooks, J. 2014. *High Tea: A Delicious Collection of Classic Afternoon Treats.* Chicago: New Hollands.

Davies, M. 2008. *American Corpus*. Retrieved August 9, 2014, from The corpus of contemporary American English COCA: http://www.americancorpus.org

Davies, M. 2010. The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English. *Language and Literary Computing*, 447-464.

Hasimoto, M. 2001. The origin of the tea plant. . *In Proceedings of 2001 International Conference on O–Cha Tea Culture and Science Session II,*, pp. 5-8.

Laybutt, B. 2009. A Cup of Tea or A Mug of Tea. *Module 6 Corpus Linguistics Course*, 1-33.

Ludeling, A., & Kyto, M. 2009. *Corpus Linguistics: An International Handbook.* Berlin and New York: Walter de Gruyter.

Paumier, S. 2014. *Unitex Manual ver 3.1.* Paris: Universite Paris Est Marne La Valee & LADL.

Pereltsvaig, A. 2016, 01 04. *Languages of the World*. Retrieved from What Will You Have Tea or Chai: http://www.languagesoftheworld.info/etymology/will-tea-chai.html

Sinclair, J. 2003. *Reading Concordance.* London: Pearson & Longman.