A CORPUS-BASED BEHAVIOURAL PROFILE STUDY OF NEAR-SYNONYMS:
IMPORTANT, ESSENTIAL, VITAL, NECESSARY AND CRUCIAL

Sani Yantandu Uba
Faculty of General Studies, Federal University Dutse, Nigeria

ABSTRACT: This study examines the internal semantic structure of a set of near-synonym adjectives: important, essential, vital, necessary and crucial. I used the British National Corpus (BNC) and 9 traditional reference materials as source data. I applied a corpus-based behavioural profile approach. The study focuses its attention on their distributional patterns especially types of nouns that each adjective modifies. Some of the findings challenge the several existing descriptions of this set of near-synonym adjectives by the traditional reference materials. For example, the traditional reference materials cite many examples of vital as modifier of abstract nouns but the corpus data indicates that it is mostly modified by concrete and dual nouns. This study justifies the theory and applicability of the BP approach for studying the semantic and usage patterns of a set of synonyms.

KEYWORDS: behavioural profile, near-synonym, distributional pattern, internal semantic structure, British National Corpus (BNC)

INTRODUCTION

Synonyms or near synonyms are complex linguistic phenomenon, which express same semantic concept in different contexts and perspectives (Liu 2010). One of its complexities is ‘two items with the same referent are not necessary synonyms since ‘synonymous’ means ‘having the same sense’ not ‘having the same referent’ (Lyons 1977). This implies that having same referent may not qualify items to be called synonym. Synonyms are often not entirely identical in meaning and hence not completely interchangeable (Liu 2010). Moreover, Divjak (2006) argues that synonyms ‘are neither in free variation, nor in complementary distribution’. Furthermore, Partington (1998) perceives synonyms as a set of lexical pairs that have similar cognitive or denotational meanings, but which may differ in collocational or prosodic behaviour. This suggests that synonyms are very challenging and complex since not all synonyms are interchangeable. They are very essential in assisting us to convey our thoughts, ideas, and feelings for effective communication (Edmonds and Hirst 2002; and Hatch and Brown, 1995). In spite of its significance a little attention has been given to the concept (Divjak 2006; Divjak and Gries 2006; Edmonds and Hirst 2002; Liu 2010; and Taylor, 2003). Consequently, more research is needed on various set of synonyms in order to shed more insight and understanding on how a set of synonyms work in terms of meanings and usage patterns. This study examines the semantic and usage patterns among the five near-synonyms adjectives: important, essential, vital, necessary and crucial, by employing a corpus-based behavioural approach. This study aims to address the following questions:

1. Does this set of near-synonyms adjectives have complex internal semantic structure?

2. Do the traditional reference materials provide adequate treatments and descriptions of this set of near-synonyms adjectives?
3. Does the corpus-based behavioural profile provide adequate treatment, descriptions and understandings of this set of near-synonyms?

The complexity of the internal semantic structure of this set of near-synonyms

One of the major reasons for studying this set of near-synonyms is its complex semantic structure and the inadequacies in the existing descriptions of the near-synonyms set. For example, all the five adjectives share same basic meanings ‘extremely important’ and ‘having a great effect’ (Cambridge Advanced Learner’s Dictionary, 2015; Longman Dictionary of Contemporary English, 2007; Oxford Advanced Learner’s Dictionary, 2010; The American Heritage Dictionary of the English Language, 2015; and Webster’s Essential Mini Dictionary 2011). Having shared same basic meanings, these adjectives appear to be interchangeable in modifying certain nouns: role and factor. For example, we can say important, crucial, necessary, essential and vital role or factor as evidenced by dictionaries. However, in some instances these adjectives cannot be interchangeable when used with certain nouns, for example, we can say crucial meeting not essential meeting, essential food not crucial food as evidenced by the British National Corpus (BNC). It is apparent that they cannot be replaced. In the light of this, it is not clear which types of nouns that these adjectives can modify, to what extent they are synonymous and which types of nouns one or some of these adjectives can be used as modifiers.

The inadequate descriptions and treatments of this set of near-synonyms by the traditional reference materials

I examined nine traditional reference materials: Cambridge Advanced Learner’s Dictionary (2015); Longman Dictionary of Contemporary English (2007); Oxford Advanced Learner’s Dictionary (2015); The American Heritage Dictionary of the English Language (2015); Webster’s Essential Mini Dictionary (2011); The American Heritage College Thesaurus (2004); The Cambridge Thesaurus of American English (1994); The Concise Oxford American Thesaurus (2006); and Merriam-Webster Dictionary of Synonym (2015). Based on the examination of this set of near-synonyms, there are inadequate descriptions of which types of nouns this set of near-synonyms modify. Moreover, there is an absence of detailed information on the usage patterns of this set of near-synonyms. These traditional reference materials offer general and core meanings of these concepts which I discussed earlier: extremely important and having a great effect. Perhaps the inadequate treatment of this set of near-synonyms is probably because of the limited space they have and the functions of these traditional reference materials perform. Liu (2010) reveals inadequate treatment of a set of near-synonyms: major, main, chief, principal, and primary. He argues that all the 12 reference materials (6 dictionaries, 4 thesauri and 2 dictionaries of synonyms) examined there were insufficient treatments of this set of near-synonyms with regard to usage patterns and which types of nouns they modify. Moreover, Church, et al. (1994) in their study of near-synonyms: ask for, demand and request discover similar insufficient detailed information on the usage patterns of this set of near-synonyms. Similarly, Tylor (2003) conducted a study on a set of near-synonyms: ‘high and tall’ and identifies inadequate treatments of these concepts by the traditional reference materials.
Testing the effectiveness of the corpus-based behavioural profile approach in examining near-synonyms set

Liu (2010) asserts that this approach of distributional patterns of lexical items was originated by Firth and his followers Halliday and Sinclair, whose work helps establish the study of lexis as a linguistic level and the examination of collocation as a key to lexical semantics in the 1950s and 1960s. Firth (1957) argues that ‘the complete meaning of a word is always contextual’ and we ‘know a word by the company it keeps’. This implies that to what extent an item is specified by its collocational environment and the tendencies of items to collocate with each other (Firth, 1966 and Sinclair, 1966).

Moreover, Sinclair (1991) states that there are two models of interpretation: the open-choice principle and the idiom principle. The former sees language as a continuous series of open-ended choices (Partington 1998) a series of slots which have to be filled from a lexicon (Sinclair 1991). This implies that open-choice principle is a way of seeing language having several complex choices and the only restraint is grammaticalness, that there are only certain items from word classes which may appear in a particular slot. While the latter emphasises that ‘the openness of choice is not available to the same extent at every point along the syntagmatic progression of an utterance’ (Partington 1998); but ‘the language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into some segments’ (Sinclair 1991). Bolinger (1976) calls these chunks of language ‘prefabrications’ or ‘prefabs’. For example, of course, set eyes on, in some cases, hard work, in spite of and so on. This implies that a language user has two options of interpretation, that of open-ended choice and idiom principle (close-ended choice). Although Partington (1998) argues that ‘the two principles work together in any text. Nor is it always easy to know, at any point in a text, which of the two principles is operating most strongly’.

Corpus-based descriptions of language have been shown to be much more accurate and informative than traditional non-corpus-based descriptions of language (Hunston, 2002 and Liu, 2010). This implies that corpus-based reference materials provide accurate and detailed information than traditional reference materials. For example, a number of corpus-based studies were conducted on pattern grammar, semantic sequences, idioms, linking adverbials, phrasal verbs and verbs (Gardner & Davies, 2007; Hunston, 2008; Hunston and Francis 1998, 2000; Liu, 2003, 2008; Moon, 1998); and a corpus-based reference grammar books were produced (Biber et al. 1999; Carter & McCarthy 2006). These reference materials had offered detailed information on the lexis and grammar usage which challenge the traditional reference materials of language descriptions.

Moreover, many scholars justify a corpus-based behavioural profile study of near-synonym is very effective (Divjak 2006; Divjak & Gries 2006; Liu, 2010; Hanks 1996; Flowerdew 2009; Gries 2001; Tylor 2003 and Walker 2011). This approach is built largely on the theory that the meaning of a lexical item correlates closely with its behavioural profile or distributional patterns (Liu 2010). Behavioural profile (BP) also refers to the description of distributional patterns of lexical items (Hanks 1996).

Firth, Halliday and Sinclair’s theoretical work on correlation between lexical semantics and distributional patterns laid the foundations for the corpus-based BP studies of lexical items (Liu 2010). He further argues that Hank’s (1996) work on corpus-based BP studies of several verbs including incite, urge, bother and abandon was the first corpus-based BPs study. The BP
Several corpus-based BPs on verbs have been established. For example, Divjak (2006) examined a set of near-synonym verbs of intending. While Divjak and Gries (2006) examined a Russian verb: trying; both studies applied a corpus-based analysis of the verbs’ morphosyntactic, syntactic, and semantic distribution patterns. Furthermore, Walker, (2011) applied a corpus-based BP study in examining a set of near-synonyms verbs: run, manage and head. He consulted two corpora: the Bank of English (BoE) and the British National Corpus (BNC). This study reveals that run is more frequently associated with nouns which refer to non-human entities such as business, companies or firms; whereas manage can be both described with nouns: non-human entities and human entities; while head, occurs more frequently with nouns which denote structures consisting of people, such as committees or panels.

The above corpus-based BPs studies are based on verbs; however, some corpus-based BPs studies on adjectives have also been established. For example, Gries (2001) conducted a corpus-based analysis of the paired English(ic and ical adjectives (economic and economical). This study offers an insight on the degree of semantic similarity and identifies difference between the two adjectives in each pair that had never been possible. Similarly, Tylor (2003) examines a set of near-synonyms: high and tall using a corpus of one million words from Lancaster-Oslo/Bergen (LOB). The results show that ‘high is by far the most frequent item. It has an overall of 579 tokens in the corpus; whereas ‘tall’ appears 59 times’. Taylor maintains that ‘both words have uses outside spatial domain, high, in particular, is widely used for non-spatial domains which are conventionally conceptualised in terms of verticality, (e.g. high temperature, high number, high pitch [of a sound] etc.’; whereas on its spatial uses ‘high’ can designate both vertical extent (high building, where the ‘height’ of the building is measured from bottom of the building up to its top) and vertical position (high ceiling, where the height of the ceiling concerns the distance from some reference point). Taylor concludes that ‘tall’ is used preferentially of human beings, trees, and other kinds of plants, as well as of buildings’; whereas high is used for all kinds of physical bodies, buildings and other kinds of constructions.

In another study, Flowerdew (2009) identifies his students (in a module on business letter writing) were not certain on ‘which adjective from a set of seemingly semantically synonymous adjectives was the right one to choose in the following sentence: thank you for your kind/sincere/cordial invitation to the alumni dinner’. Flowerdew consulted a Business Letters Corpus in order to search these different expressions. The results indicate that cordial + invitation is used for offering an invitation (e.g. may we extend to you a cordial invitation); whereas kind + invitation is used for accepting an invitation or thanking the host (e.g. thank you very much for your kind invitation). Liu (2010) examines five near-synonym adjectives: chief, main, major, primary, and principal by applying a corpus-based BPs approach. The major findings of this study are: (i) chief and principal are not frequently used to modify abstract/dual nouns and, when they do, they are used mostly in formal registers such as academic writing; (ii) principal is used mostly in non-position titles to mean highest degree of contribution or share rather than authority or power; (iii) major and primary can be used as predicative adjectives; and (iv) the meanings of the adjectives, especially in the case of
primary, are context-dependent. This corpus-based BPs study has also yielded some interesting findings that challenge some of the existing descriptions of these adjectives in traditional reference materials.

The above literature review indicates that corpus-based BPs studies are effective approaches to understanding near-synonyms. Subsequently, I will apply this approach in order to identify usage and semantic differences among the five near-synonym adjectives.

METHODOLOGY

In this study, I applied a corpus-based analysis; and used the British National Corpus (BNC) of a 100-million-word and nine traditional reference materials as source data. The nine traditional reference materials are: 5 dictionaries, Cambridge Advanced Learner’s Dictionary (2015); Longman Dictionary of Contemporary English, (2007); Oxford Advanced Learner’s Dictionary (2015); The American Heritage Dictionary of the English Language (2015); Webster’s Essential Mini Dictionary (2011); three thesauri, The American Heritage College Thesaurus (2004); The Cambridge Thesaurus of American English (1994); The Concise Oxford American Thesaurus (2006); and one dictionary of synonym, Merriam-Webster Dictionary of Synonym (2015). The rational for using the BNC is on the premise that it is a reference corpus because it is the largest English reference corpus.

EVIDENCE FROM THE CORPUS: DATA ANALYSIS AND DISCUSSION

The overall usage patterns

The overall usage pattern is one of the foci that may offer invaluable insights and understandings on how this set of near-synonyms function. One way to understand the overall usage pattern is to get the overall frequency of this set of near-synonyms. I used part of speech query of each adjective from the search engine of the BNC in order to get the overall frequency. The results are in table 1 below.

Table 1. Frequency of the five set of near-synonym adjectives

<table>
<thead>
<tr>
<th>Crucial</th>
<th>Vital</th>
<th>Essential</th>
<th>Necessary</th>
<th>Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>4,402</td>
<td>5,035</td>
<td>8,633</td>
<td>17,874</td>
<td>38,716</td>
</tr>
</tbody>
</table>

This table shows the frequency of these adjectives from the lowest to the highest. Although the results do not indicate any usage patterns but it shed more light and insight on their frequency distribution. This information offers significant variations of frequency distribution among them which in turn may affect their usage patterns. For example, crucial has frequency of 4,402 and important has 38,716. However, based on the dictionaries and thesauri consulted these five adjectives appear to have core meanings of ‘having a great effect’ and ‘extremely important’ as discussed earlier’ but their frequencies differ significantly. This may probably provide some similarities or differences in their usage patterns. This brings the issue of identifying which types of nouns they modify. Do they modify same types of nouns, do they modify different types of nouns or do they overlap in modifying all or some types of nouns.
To identify which types of nouns each adjective modify, I queried the BNC via part of speech tag; which facilitated generation of the most frequent collocates of this set of near-synonym adjectives. The results show nouns which are most frequently modified by each adjective.

**Table 2. Top ten nouns most frequently modified by each of the five adjectives**

<table>
<thead>
<tr>
<th>S/NO.</th>
<th>Crucial</th>
<th>Vital</th>
<th>Essential</th>
<th>Necessary</th>
<th>Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Role</td>
<td>Role</td>
<td>Oils</td>
<td>Condition</td>
<td>Factor</td>
</tr>
<tr>
<td>2</td>
<td>Factor</td>
<td>Ingredient</td>
<td>Ingredient</td>
<td>Skill</td>
<td>Role</td>
</tr>
<tr>
<td>3</td>
<td>Distinction</td>
<td>Importance</td>
<td>Element</td>
<td>Step</td>
<td>Part</td>
</tr>
<tr>
<td>4</td>
<td>Element</td>
<td>Organ</td>
<td>Feature</td>
<td>Longer</td>
<td>Play</td>
</tr>
<tr>
<td>5</td>
<td>Difference</td>
<td>Clue</td>
<td>Component</td>
<td>Information</td>
<td>Issue</td>
</tr>
<tr>
<td>6</td>
<td>Play</td>
<td>Play</td>
<td>Part</td>
<td>Action</td>
<td>Aspect</td>
</tr>
<tr>
<td>7</td>
<td>Issue</td>
<td>Supplies</td>
<td>Characteristic</td>
<td>Measure</td>
<td>Thing</td>
</tr>
<tr>
<td>8</td>
<td>Understanding</td>
<td>Part</td>
<td>Oil</td>
<td>Adjustment</td>
<td>Point</td>
</tr>
<tr>
<td>9</td>
<td>Success</td>
<td>Element</td>
<td>Tool</td>
<td>Resource</td>
<td>Feature</td>
</tr>
<tr>
<td>10</td>
<td>Determinant</td>
<td>Factor</td>
<td>Prerequisite</td>
<td>Arrangement</td>
<td>element</td>
</tr>
</tbody>
</table>

This result indicates the most top ten nouns that each of the five adjectives modify. Although there are some overlaps but the result offers significant insights.

To identify the differences among the types of nouns they modify, I adopted Liu’s (2010) category of nouns into three semantically groups: abstract (e.g. issue, success) concrete (e.g. tool, oil) and dual (e.g. action, role). The results of the classification are presented in table 3.

**Table 3. Types of nouns modified by each adjective**

<table>
<thead>
<tr>
<th>Crucial</th>
<th>Vital</th>
<th>Essential</th>
<th>Necessary</th>
<th>Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 abstract</td>
<td>2 abstract</td>
<td>3 abstract</td>
<td>7 abstract</td>
<td>6 abstract</td>
</tr>
<tr>
<td>1 concrete</td>
<td>4 concrete</td>
<td>4 concrete</td>
<td>1 concrete</td>
<td>1 concrete</td>
</tr>
<tr>
<td>2 dual</td>
<td>4 dual</td>
<td>3 dual</td>
<td>2 dual</td>
<td>3 dual</td>
</tr>
</tbody>
</table>

The results show types of nouns that each adjective modifies among the top ten nouns. It offers significant information on which types of nouns that each adjective modifies. For example, crucial and necessary out of the top ten nouns they modify each has seven abstract nouns; whereas important has six out of ten. Moreover, these three adjectives have one concrete noun each. In contrast, vital and essential have four concrete nouns each; and four dual for vital while three dual for essential; and they have less than three for abstract nouns. This revelation indicates that crucial, necessary and important are mostly modified abstract nouns; whereas vital and essential are mostly modified concrete and dual nouns.

One important revelation of this result is vital which predominantly modifies concrete and dual nouns. In contrast, the traditional reference materials I consulted cited numerous examples of abstract nouns. For example, ‘the work she does is absolutely vital’ (Longman Dictionary of Contemporary English, 2007); ‘it is vital you keep accurate records when you are self-employed’ (Oxford Advanced Learner’s Dictionary, 2010); ‘a strong opposition is vital to a healthy democracy’ (Cambridge Advanced Learner’s Dictionary, 2015) and ‘tourism is vital to the country’s economy’ (Webster’s Essential Mini Dictionary, 2011). This result challenges
the significant use of abstract nouns by these dictionaries to modify *vital* whereas the corpus data shows abstract noun is rarely used to modify *vital*.

Furthermore, the core meanings of these adjectives as discussed earlier are ‘*extremely important*’ and ‘*having a great effect*’. These meanings suggest that these adjectives are interchangeable but the corpus data shows contrary. For example, I queried the BNC on this phrase ‘*crucial food*’, there was no any citation of this phrase in the BNC; and I also made similar query on ‘*essential meeting*’ but there was no any citation in the BNC. This suggests that even though these adjectives have same core meanings there are some instances which cannot be interchangeable. This finding corroborates Liu’s (2010) claim that synonyms are often not entirely identical in meaning and hence not completely interchangeable.

Based on the corpus-based BP studies of this set of near-synonym adjectives, this study has identified some interesting findings which challenge the existing descriptions of these adjectives by the traditional reference materials. It has also made a major contribution towards addressing some important theoretical issues that the paper sets to address. The first issue is testing and exploring the theory and applicability of the corpus-based behavioural profile approach to the study of synonymous adjectives. The second aspect is contributing to the development of more adequate descriptions of the near-synonyms. Thirdly, there are inadequate treatments of this set of synonymous adjectives by the traditional reference materials.

In terms of testing and exploring the theory and applicability of the corpus-based BP approach, this study reveals interesting findings. For example, the corpus-based BP has provides useful information and description of this set of adjectives which shows that the approach is appropriate in examining near-synonym. The close examination of semantic nouns that each adjective modifies and their collocational patterns has resulted to the accurate identification of the distinctive semantic patterns of this set of adjectives. This study corroborates similar studies (Gries, 2001 and Liu, 2010) which identify types of nouns that adjectives modify and the accompanying syntactic features of these adjectives. The examination of semantic occurents of near-synonyms may vary from study to study. For example, studies of the near-synonymous verbs of Tylor (2003); Divjak & Gries (2006) and Walker (2011) applied more advanced data and more occurents; while this study like Hanks (1996) and Liu (2010) examine few co-occurrants and their accompanying syntactic properties.

With regard to the contribution to the development of adequate descriptions of the near-synonyms, this study discovers that the coverage of this set of near-synonyms in the traditional reference materials is inadequate in some instances. This inadequate treatment in the traditional reference materials corroborates Liu (2010) findings of inadequate description of near-synonym set. These findings can be used by the reference material writers to improve and address inadequate descriptions of their reference materials. Similarly, useful and insightful information were acquired in this study that crucial, *necessary* and *important* are mostly modified abstract nouns; whereas *vital* and *essential* are mostly modified concrete and dual nouns. Furthermore, it reveals that while all the dictionaries consulted cited more examples of *vital* with abstract nouns but the corpus data indicates that the adjective in question is mostly modified by concrete and dual nouns.
CONCLUSION

In sum, this study provides invaluable insights and descriptions of internal semantic structures of this set of synonyms, which highlights the inadequate treatments of these adjectives by the traditional reference materials. Thus, applying a corpus-based BPs will provide detailed information and description of internal semantic structure of near-synonyms.

FURTHER RESEARCH

This research is concerned with single set of near-synonyms of adjectives, a study on multiple sets of near-synonyms is needed which will provide further insights and descriptions of more near-synonyms sets.

REFERENCES

British National Corpus (BNC) accessed via www.sketchengine.co.uk on 14th May, 2015


