

## A Corpus-based Study of Similes in British and American English

**Riyad F Hussein**

Department of Linguistics, Faculty of Foreign Languages  
The University of Jordan  
Amman, Jordan

**Majdi Sawalha**

Computer Information Systems Department  
King Abdullah II School for Information Technology  
The University of Jordan  
Amman, Jordan

### Abstract

This study aimed to investigate the forms of similes and the types of nouns following them and their frequency in different simile constructions. It also unveils the implications that can be drawn from the findings for English as Foreign Language (EFL) learning contexts. More specifically, this study attempts to find out the most frequent nouns following similes in different constructions in two generalized corpora, namely the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). These similes were identified, retrieved and ranked in a descending order according to their frequencies per one million words. Results indicate that similes, unlike other multi-word units are changeable as different nouns with varying frequencies ranging from one to fifty three were shown to follow each of the similes under investigation. For instance, frequencies of nouns following the simile *as good as* were as follows: *gold* (53), *(the) people* (30), *men* (14), *money* (9), and *cash* (8). These results stood in contrast to the beliefs and opinions of some native speakers of English who view similes as fixed forms, very much like formulaic expressions and thus associated exclusively with certain nouns. The study calls on teachers and instructors to take these findings into account when teaching similes in the EFL context. Additional research is recommended on other similes such as, *as sweet as sugar* and *as cold as ice*, for instance in BNC and COCA to confirm or invalidate the findings reported in this research.

**Keywords:** BNC, COCA, corpora, corpus linguistics, EFL, similes

**Cite as:** Hussein, F. R., & Sawalha. M . (2016). A Corpus-based Study of Similes in British and American English. *Arab World English Journal*, 7(2). DOI: <https://dx.doi.org/10.24093/awej/vol7no2.4>

## Introduction

A simile is a figure of speech that compares two different things or people to identify similarities or qualities which are shared by them by using the word "like" or "as" to make the comparison. Similes are generally more easily identified because of the use of the word "as" and/or *like* and metaphors are figures of speech used mainly in the analysis of literary texts. Whereas the former compares things which share a common feature, the latter does so more directly without using "like" or "as", as shown in the sentence: *He is a lion*. The impression among some students of literary expression and foreign language learners is that similes can be expressed through only one frame namely, [As+Adj+As+Noun], but the fact is that similes can be expressed through the use of other frames such as [Adj+like+Noun] as in "*hard like brick*" and [Adj+as+Noun] as in "*black as coal*."

When similes are taught in English as a foreign language (EFL) classes or in the analysis of literary texts at the college level, students are instructed explicitly to learn similes in much the same way as formulaic expressions that is in forms which are fixed and not changeable such as *by the way*, *so far so good*, *see you later* and *by all means* because for instance when the simile *as old as the hills*, it is usually emphasized that the form "the hills" was the only correct form and any other noun replacing "the hills" was considered incorrect. This, despite the fact that native speakers of English sometimes use other nouns to replace the noun *the hills*. In brief, it seems that similes are unlike idioms because idioms are culture-bound and fixed expressions are unchangeable. Similes, however, are changeable across language varieties and dialects, but this is subject to verification by the analysis of the data that will be obtained in this study.

## Corpora and Language analysis

Corpora are important tools for the study of language. In the last forty or fifty years and with the availability and increasing use of electronic corpora, their role is getting increasingly important in the study of language variants, forms, and styles. Although no one can underestimate the role of native English speakers in identifying the correct grammatical, lexical or stylistic forms, lack of consensus among them as to the most acceptable or correct form makes their judgments sometimes less reliable. Therefore, native speakers cannot always be taken as undisputable arbiters on language issues such as the most frequent words in English, or differences between spoken and written English, tenses that people use more frequently and idioms or similes people frequently use. This perhaps explains why linguists, grammarians, researchers and dictionary compilers resort to corpora to verify the correct language forms and to find answers to controversial questions or issues. McCarthy (2004) confirms that:

With corpora and software tools to analyze them, we can see how language is really used. We no longer have to rely heavily on intuition to know what we say or what we write; instead we can see what hundreds of different speakers and writers have actually said or written, all at the click of a mouse (p. 2)

That is why language researchers, linguists and grammarians resort to corpora, either generalized or specialized depending on the goals and scope of their research. Bennett, (2010) states "Generalized corpora are often very large, more than 10 million words and contain a variety of language so that findings from it can be somewhat generalized" (p. 13). There are many generalized corpora such as The British National Corpus (BNC), and the American

National Corpus (ANC) and the Corpus of Contemporary American English (COCA) which contains approximately 450 million words. The generalized corpora contain written texts such as newspaper and magazine articles and spoken transcripts such as conversations and business meetings. Large, generalized corpora should be used if valid generalizations about language are to be drawn. Unlike generalized corpora, specialized corpora are often created to address very specific questions. The most widely used specialized corpora are The Michigan Corpus of Academic Spoken English (MICASE) and Corpus of Upper Level Student Papers (MICUSP) which contains papers written by advanced students from a range of university disciplines such as English, education, engineering, philosophy, political science, psychology and history.

### **Intuition in corpus linguistics**

Intuition in corpus linguistics is a controversial issue because whereas some scholars approve and subscribe to the use of intuition due to its value in formulating research questions and conceptualizing areas of research topics, others doubt the validity of intuition particularly in the analysis of speech. A compromise approach however is in favor of reliance but not overreliance on intuition.

Gries (2006) maintains that “intuitive/subjective decisions come into play at different points of time”. (p. 86). According to him, intuition is involved in the identification of a topic or problem and to a lesser extent in the retrieval of data where a decision for the selection of a particular corpus is to be made and finally during the coding process and statistical analysis.

Conrad (2006) gives some credit to the use of intuition in corpus linguistics. She believes that intuition can play a role in helping analysts develop research questions. She notes, “your intuitions about how you use language can lead to many useful corpus investigations, as can disagreements in speakers' intuitions”. (p. 51). Finally, Hyland (2006) is in favor of reliance on intuition as we have intuitions about what to look for and then interpret what things mean. He is, however, against over-reliance on intuition. He argues:

Obviously it is the over-reliance on intuition that has attracted criticism in the past –cases where sometimes whole theories of language were based on armchair theorizing and invented examples. But while intuition is generally a poor guide to judgments about frequency, collocation, semantic meanings, phraseology etcetera, interpreting is important when generalizing from corpus data and understanding the numbers and patterns we find in it (p. 103).

Finally, Aston (2006) doubts the validity of intuition which is particularly unreliable for speech. He maintains:

Our intuition tends to focus on one use at a time, forgetting others than the one we first thought of –particularly uses in other text –types and context types. Intuition is also notoriously unreliable as to the relative frequency of different features, and obviously, is hopeless with regard to text –and context types with which we are not familiar (p. 5)

In investigating language forms and language use, generative linguists relied heavily on the intuition of native speakers which assumes an 'ideal native speaker/hearer' who speaks an invariant variety of the language in question. But Leech (2006) maintains that,

Sociolinguists and other usage –oriented branches of linguistics have highlighted the variability of the competences of different native speaker dialects or idiolects. As the non-uniformity of a language is widely accepted as self-evident, it is clear that the native speaker's knowledge of that language is bound to be incomplete, whether in terms of register, dialect or diachrony (p. 162)

In addition, native speaker's intuition varies considerably from that of other native speakers due to experience, language creativity, knowledge of other languages and educational level, as to the acceptability of forms and their meanings, and so the form that is acceptable by one native speaker may not be acceptable by others. It is not unusual that when a question or a language problem is raised by EFL learners, native speakers come up with diverse answers and express divergent opinions. This explains perhaps why some English grammar books are written by nonnative speakers with or without the collaboration of native speakers. A case in point is, *A Practical English Grammar* by Thomson and Martinet (1986) and, *Communicative Grammar of English*, by Leech and Svartvik (2013). The lack of consensus among native speakers is due to the variation from one genre to another, and from one dialect to another and diachronic variation that exists over time.

If intuition is partly reliable and if native speakers cannot consistently offer substantial help in determining the correct or acceptable forms, then the assistance of another source be sought, namely language corpora. Corpora can systematically provide answers to language questions or issues and queries such as word counts word frequency, forms and their co-occurrences with other word forms as collocations and colligations, multi-units, use and frequency idioms, formulaic expressions and similes across different registers, lexicography and more specifically lexico-grammatical patterns of language as performed by Sinclair and his Cobuild project team (Sinclair 1991). This in addition to research on register variation conducted by Biber (1988) and the difference between features of conversation and written discourse as explained by Carter and McCarthy (1995), and language use whether it is inherently variable or heterogeneous as stated by Labov (1972) or systematically heterogeneous as stated by Halliday (1991), is then a way out for language researchers, grammarians, lexicographers, dictionary writers and foreign language learners through the use of electronic corpora both generalized and specialized depending on areas or topic to be investigated.

The paper will be presented in the following order, section four provides a brief summary of the methods and objectives of this research, while section five presents the data analysis, and the last section presents the summary and conclusion.

### **Methodology and Objectives**

The data collection process was carried out through the use of two corpora, the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). Four randomly selected simile forms were investigated, involving the adjectives, *good*, *old*, *white* and *black* in reference to the constructions [As+Adj+As+Def.Art.+N], [As+Adj+As+Indef.Art.+N]

and [As+Adj+As+0Art.+N] in both the BNC and COCA. The four similes in their different constructions were retrieved, examined, and ranked in a descending order according to their frequency in BNC. Then they were described and analyzed.

One reason why we embarked on the investigation of similes is due to the fact that similes are believed to be fixed and not changeable in form in the context of English as a foreign or native language. In line with this conceptualization, when similes are taught in English as a foreign language classes or at college level, students are generally given the impression that similes are fixed, multi-unit words, very much like idioms. They are taught, for instance, that with the simile, *as white as snow*, the only correct noun following it is *snow*. But streaming the corpora certainly suggests this is not the case because one may come across other nouns following this simile in both the BNC and COCA.

To retrieve data from corpora, it is essential to use specific language software programs or tools. Aasheim (2012) asserts that, “A corpus often has as an inbuilt search engine, which makes it possible for search for different words or constructions in order to test hypotheses, check occurrences and validate rules of usage” (p. 6). Each corpus generally has a software or an interface which is very much like a key through which one can have access to publicly available corpora or restricted corpora. To obtain data from the British National Corpus one has to do so through Brigham Young University Interface, (BNC, BYU). The objective of this study is to find out the similarities and/or differences in the different constructions of similes in BNC and COCA, and attempts to highlight the implications which can be drawn from the findings for EFL learning contexts. More specifically this study will address the following questions:

- 1) What are the most frequent nouns following the simile frame [As good as...] in its different constructions in both BNC and COCA?
- 2) What are the most frequent nouns following the simile frame [As old as...] in its different constructions in both BNC and COCA?
- 3) What are the most frequent nouns following the simile frame [As black as...] in its different constructions in both BNC and COCA?
- 4) What are the most frequent nouns following the simile frame [As white as...] in its different constructions in both BNC and COCA?

### Results related to the first question

In addressing the first question, it is clear that the most frequent nouns following the simile construction [As+ good+ as+ Def.Art.+ Noun] in BNC are *the people* (4), *the food* (2), *the state*(2), *the quality* (2), and their normalized frequency (occurrences per million words) in BNC are 0.04, 0.02, 0.02, and 0.02 respectively, whereas in COCA they are *the people* (17), *the day* (4), *the quality* (3), *the originals* (3), *the men* (3), and *the Japanese* (3), and their normalized frequency (occurrences per million words) are 0.04, 0.01, 0.01,0.01, 0.01 and 0.01 respectively and ratio refers to the relative percentage in the two corpora as shown in Table 1. It is also clear that the nouns *the people* (21), *the quality* (5), and *the day* (5) ranked highest in both corpora as shown in Table 1.

**Table 1** *The most frequent nouns following the simile “as good as the” in BNC and COCA*

WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
-------------	--------	---------	------	------	-------

1	As good as the people	4	17	0.04	0.04	1.06
2	As good as the food	2	2	0.02	0.00	4.50
3	As good as the state	2	0	0.02	0.00	2.00
4	As good as the quality	2	3	0.02	0.01	3.00
5	As good as the day	1	4	0.01	0.01	1.13
6	As good the originals	1	3	0.01	0.01	1.50
7	As good as the men	1	3	0.01	0.01	1.50
8	As good as the Japanese	1	3	0.01	0.01	1.50
9	As good as the man	1	2	0.01	0.00	2.25
10	As good as the stage	1	2	0.01	0.00	2.25

With regard to the most frequent nouns used in the construction [As+ good+ as+ Indef.Art.+ Noun], they are *arest* (5), *a man* (2), *a play* (2), and *a wink* (2) in BNC and *a man* (5), *a rest* (1), *a play* (1) and *a wink* (1) in COCA as shown in Table 2.

**Table 2** The most frequent nouns following the simile “as good as a(an)” in BNC and COCA

No	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As good as a rest	5	1	0.05	0.00	22.50
2	As good as a man	2	5	0.02	0.01	1.80
3	As good as a play	2	1	0.05	0.00	9.00
4	As good as a wink	2	1	0.02	0.00	9.00

Table 3 shows that the most frequent nouns following the simile construction [As+ good+ as+ 0.Art.+ Noun] in BNC are *gold* (20), *cash* (4), *people* (3) and *men* (3), whereas the most frequent nouns in COCA are *gold* (33), *men* (8), *money* (7), *people* (6) and *cash* (4).

**Table 3** The most frequent nouns following the simile “as good as +0.Art.” in BNC and COCA

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As good as gold	20	33	0.20	0.07	2.73
2	As good as cash	4	4	0.04	0.01	4.50
3	As good as people	3	6	0.03	0.01	2.25
4	As good as men	3	8	0.03	0.02	1.69
5	As good as money	2	7	0.02	0.02	1.29
6	As good as chance	2	3	0.02	0.01	3.00

The last three tables clearly show the most frequent nouns following the simile [As + good + as+Noun] in its various constructions in both BNC and COCA are *gold* (53), *people* or *the people* (30), *men* (14), *money* (9), *cash* (8) and *a man* (7).

**Results related to the second question**

Table 4 shows that the most frequent nouns following the simile construction [As+ old + as+ Def.Art+ Noun] are *the hills* (13), *the history* (3), and *the world* (2) in BNC, whereas in COCA they are *the hills* (18), *the world* (7), *the history* (2) and *the concept* (2). The most frequent nouns in both corpora are *the hills* (31), *the world* (9) and *the history* (5).

**Table 4** *The most frequent nouns used with the simile “as old as the” in BNC and COCA*

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As old as the hills	13	18	0.13	0.04	3.25
2	As old as the history	3	2	0.03	0.00	6.75
3	As old as the world	2	7	0.02	0.02	1.29
4	As old as the concept	1	2	0.01	0.00	2.25
5	As old as the man	1	1	0.01	0.00	4.50
6	As old as the game	1	1	0.01	0.00	4.50

With regard to the most frequent nouns used in the construction [As+ old+ as+ Indef. Art.+ Noun], they are *a hat* (1), and *a Pharoah* (1) with a very low frequency in BNC, and with a zero frequency in COCA as shown in Table 5.

**Table 5** *The most frequent nouns following the simile “as old as a(an)” in BNC and COCA*

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As old a hat	1	0	0.01	0.00	1.00
2	As old as a Pharoah.	1	0	0.01	0.00	1.00

The most frequent nouns following the simile construction [As+ old+ as+ 0.Art.+ Noun] are *time* (4), *history* (3), *human* (2) and *America* (2) in BNC, and *time* (27), *man* (7), *history* (6), *human* (6) and *America* (5) in COCA as shown in Table 6.

**Table 6** *The most frequent nouns following the simile “as old as + 0.Art” in BNC and COCA*

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As old as time	4	27	0.13	0.04	3.25
2	As old as history	3	6	0.03	0.00	6.75
3	As old as human	2	6	0.02	0.02	1.29
4	As old as America	2	5	0.01	0.00	2.25
5	As old as man	1	7	0.01	0.00	4.50
6	As old as humankind	1	3	0.01	0.00	4.50

Tables 4, 5, and 6 clearly show that the most frequent nouns following the simile [as old as+Noun] in both BNC and COCA are *the hills*(31), *time* (31), *(the) history* (14), *(the) man* (10), *the world* (9), *human* (8) and *America* (7).

**Results related to the third question**

The most frequent nouns following the simile construction [As+ white+ as+ Def.Art.+ Noun] in BNC are *the snow* (1), *the towel* (1), and *the wall*(1) in BNC, whereas in COCA it is *the snow* with a frequency of (8) as shown in Table 7.

**Table 7** The most frequent nouns following the simile “as white as the” in BNC and COCA.

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As white as the snow	1	8	0.01	0.02	0.56
2	As white as the towel	1	0	0.01	0.00	1.00
3	As white as the wall	1	0	0.01	0.00	1.00

With regard to the most frequent nouns following the construction [As+ white+ as+ Indef. Art.+ Noun], they are *a sheet* (13) and *a ghost* (3) in BNC and the nouns *a sheet* (9) and *a ghost* (6) in COCA as shown in Table 8.

**Table 8** The most frequent nouns following the simile “as white as a(an)” in BNC and COCA

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As white as a sheet	13	9	0.13	0.02	6.50
2	As white as a ghost	3	6	0.03	0.01	2.25

Table 9 shows that the most frequent nouns following the simile construction [As+ white+ as+ 0.Art.+ Noun] are *snow* (9) and *paper* (2) in BNC, and *snow* (14), *chalk* (7), and *milk* (5) in COCA.

**Table 9** The most frequent nouns following the simile “as white as+0.Art.” in BNC and COCA

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As white as snow	9	14	0.09	0.03	2.89
2	As white as paper	2	1	0.02	0.00	9.00
3	As white as chalk	1	7	0.01	0.02	0.64
4	As white as milk	1	5	0.01	0.01	0.90

Tables 7, 8, and 9 clearly show that the most frequent nouns following the simile [as white as +Noun] in both BNC and COCA are (*the*) *snow* (32), *a sheet* (22), *a ghost* (9), *chalk* (8) and finally *milk* (6).

**Results related to the fourth question**

Table 10 shows that the most frequent nouns following the simile [As+ black + as+ Def.Art.+ Noun] in BNC and COCA are *the night* (3), *the ace* (2), and *the devil* (2).

**Table 10** The most frequent nouns following the simile “as black as the” in BNC and COCA

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As black the night	1	2	0.01	0.00	2.25
2	As black as the ace	1	1	0.01	0.00	4.50
3	As black as the devil	1	1	0.01	0.00	4.50

The most frequent noun following the simile construction [As + black + as + Indef. Art.+ Noun] in BNC and COCA is *raven* (2) as shown in Table 11.

**Table 11** The most frequent nouns following the simile “as black as a(n)” in BNC and COCA

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As black as a raven	1	1	0.01	0.00	4.50
2	As black as a stone	1	0	0.01	0.00	1.00
3	As black s storm	1	0	0.01	0.00	1.00
4	As black as a Luger	1	0	0.01	0.00	1.00

The most frequent nouns following the simile construction [As + black + as+0.Art+Noun] are *thunder* (3), *midnight* (2), and *hell*(2) in BNC and *coal* (6), *night* (3) and *midnight* (2) in COCA as shown in Table 12.

**Table12** The most frequent nouns following the simile “as black as+0.Art” in BNC and COCA

	WORD/PHRASE	1: BNC	2: COCA	PM 1	PM 2	RATIO
1	As black as thunder	3	0	0.03	0.00	3.00
2	As black as midnight	2	2	0.02	0.00	4.50
3	As black as hell	2	0	0.02	0.00	2.00
4	As black as coal	1	6	0.01	0.00	1.00
5	As black as night	1	3	0.01	0.01	1.50
6	As black as ebony	1	1	0.01	0.00	4.50

Tables 10, 11, and 12 clearly show the most frequent nouns following the simile [As black as+ Noun] in both BNC and COCA are *(the) night* (7), *coal* (7), *midnight* (4) and *thunder* (3).

### Summary and Conclusion

In this study both quantitative and qualitative approaches were used to understand and analyze the findings and quantitative results generated from the corpus were analyzed qualitatively. Qualitatively analyzing results involved examining the simile forms, the most frequent nouns following the simile construction in both BNC and COCA, and interpreting and explaining them.

Analysis of the data suggests that there are similarities as to the nouns following the different simile constructions in both the BNC and COCA corpora but with different frequencies. For instance, the most frequent nouns following the simile “as old as” in BNC were *the hills*

(13), *(the) history* (6), *time* (4), *human* (2) and *America* (2) whereas in COCA they were *time* (27), *the hills* (18), *(the) history* (8), *man* (7), *the world* (7), *human* (6) and *America* (5). However, the most frequent nouns following the simile “*as old as+Noun*” in both BNC and COCA were *the hills* (31), *time* (31), *(the) history* (14), *(man) or the man* (10), *the world* (9), *human* (8) and *America* (7). Regardless of the high or low frequency of nouns following simile constructions, what is noteworthy here is the number of multiple nouns that can follow a simile form. This shows beyond doubt that there is some but not total consensus among native speakers of English as to what noun should follow the simile “*as old as*” or other similes for that matter. If the BNC faithfully represents British English use and if COCA faithfully represents American English use, then it can be concluded that native speakers of English opt for the use of the nouns *the hills* and *time* more than the nouns *the world* or *America* or *the concept* or *the game* in the simile “*as old as*”. But this of course suggests that it is possible to say *as old as America* or *as old as the world* in much the same way we say *as old as the hills* or *as old as time* which are used but not so frequently by native speakers. What determines the form of a simile seems to be directly related to the locale or geographical area. So in a very cold area such as ice-cold Minnesota, it would not be surprising if people say *as white as snow*, but in an area well-known for milk production and dairy farming, people may opt for the simile *as white as milk*.

When similes are taught in the foreign language context or introduced in English literature classes or used in the analysis of literary texts at the college level, students are instructed explicitly to learn similes and use them in much the same way as formulaic expressions, that is in forms which are fixed and not changeable such as, *by the way*, *so far so good*, *beg your pardon*, *by all means* and *see you later*. It may not therefore be premature to say that similes like *as hard as a rock*, *as busy as a bee*, and *as happy as a lark* are not the only acceptable forms in English because they can be rendered in other forms such as, *as hard as iron*, *as busy as a beaver* or *an ant* and *as happy as a clam*. But this of course needs further investigation to confirm the types of nouns following the simile constructions and their frequency as well. In brief, it seems that similes are unlike idioms because idioms are culture-bound and fixed expressions and are unchangeable whereas similes are changeable across language varieties and dialects, so similes in the US may be different in form from those used in New Zealand and in turn those used in New Zealand may be different from those used in the United Kingdom.

The implications that can be drawn from the findings of the study for the EFL learning context is that there may not be only one correct simile form which is given priority over others but rather a number of correct and perhaps equally acceptable forms. Because more than one noun can follow a simile in native English, we believe EFL learners should also be given the opportunity to release their imaginative powers to come up with new simile forms based on their experience, not only in the target language but also in their native language and culture as is the case with native speakers of English who for instance in the BNC were shown to use approximately forty nouns each of which with a frequency of only one following the simile “*as good as..*”; some of these nouns are, *the car*, *the king*, *the lab*, *the team*, *the tools*, *the lake*, *the woman*, *the Vauxhall*, *the dinners*, *the banana* and *the rest*, in addition to *a signature*, *a ride*, *a pair*, *a day*, *a fingerprint*, *a master*, *a murderer*, *a ballet*, *a caress*, *a cuddle*, *a pauper*, *a seal*, *a frolic*, *a giggle*, *a holiday*, *a house*, *a lead*, *a doctor*, *a dozen*, *a slide*, *a squirrel*, *a streetcar*, *a tonic*, *a week-end*, and *a moon*, and *parties*, *partnership*, *master*, *inputs*, *Nescafe*, *Duke*,

*foundation*. In COCA there were also ten nouns, namely, *image, individuals, the input, the music, the players, the price, the pros, the stones, the sum, and the women* following the simile “as good as” and each of which with a frequency of only one.

In addition, in the BNC there were approximately six nouns each of which with a frequency of only one following the simile “as white as”; these nouns are *wedding, Cleopatra, death, ivory, marble, and office*.

To reiterate, there seems to be no consensus as there is a variation as to what noun should follow a certain simile as shown by the data obtained from the BNC and COCA; it is suggested therefore that this liberal tendency towards the use of similes in native English be extended to the context of English as a foreign language so that students can enrich the language by their creative, innovative and idiosyncratic forming of similes based on the totality of their experience in language and culture.

#### About the authors:

**Riyad F Hussein** is a Professor in the department of linguistics at the University of Jordan, Amman, Jordan. He obtained his Ph.D in linguistics from the State University of New York (SUNY) at Buffalo. He has published extensively in international journals such as *Language Problems and Language Planning, English for Special Purposes, Language Sciences, World Englishes, Babel and META*. His research interests include bilingualism, code-switching, second language acquisition, translation and interpreting

**Majdi S Sawalha** is an assistant professor in the Computer and Information Systems Department, The University of Jordan. He obtained his Ph.D from the university of Leeds in 2012. His research interests include Computational Linguistics, Corpus Linguistics, Morphological and Syntactic Analysis, Semantic Analysis, Prosodic and Phonological Analysis, Text Data Mining, Lexicography and Machine Learning. He has published numerous articles in regional and international journals. He has also developed the SALMA toolkit for morpho-syntactic analysis of Classical and Modern Standard Arabic.

#### References

- Aasheim, I. (2012). *A Contrastive Study of Similes in English and Norwegian*. Unpublished Ph.D dissertation, Oslo: University of Oslo
- Aston, G. (2006). Applied Corpus Linguistics and the Learning Experience In V. Viana, S. Zyngier, & G. Barnbrook, (eds.) *Perspectives on Corpus Linguistics* (pp.3-16). Amsterdam: John Benjamins Publishing Company
- Benett, Gena R. (2010). *Using Corpora in Language Learning Classroom: Corpus Linguistics for Teachers*. Ann Arbor: University of Michigan Press
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press
- Carter, R. & McCarthy, M. (1995) Grammar and the Spoken language in *Applied Linguistics*, 16:141-158
- Conrad, S. (2006). Variation in corpora and its pedagogical implications In V. Viana, S. Zyngier, & G. Barnbrook (eds.) *Perspectives on Corpus Linguistics*. (pp.47-62). Amsterdam: John Benjamins Publishing Company

- Corpus of Contemporary American English*. Retrieved February 23, 2015 from <http://corpus.byu.edu/bnc/>
- Durand, J. (2009) On the Scope of Linguistics: Data, Intuition, Corpora in Y. Kawaguchi, M. Mingshi, & J. Durand . (eds.) *Corpus Analysis and Variation in Linguistics*. (pp.25-52) . Tokyo: Tokyo University of Foreign Studies.
- Gries, Stefan Th. (2006). Methodological and Interdisciplinary Stance in Corpus Linguistics In V. Viana, S. Zyngier, & G. Barnbrook, (eds.) *Perspectives on Corpus Linguistics* (pp.81-98). Amsterdam: John Benjamins Publishing Company
- Halliday, M.A.K. (1991). Corpus Studies and Probabilistic Grammar in English Corpus Linguistics In B. Aijmer & Altenberg (eds.). *Advances in Corpus-Based Contrastive Linguistics: studies in honor of Stig Johansson* London: Longman
- Hasselgård, H. Johansson, S. & Lysvag, P (1998). *English Grammar: theory and use*. Oslo: Universitetsforlaget
- Hyland, K. (2006) Looking through corpora into writing practices In V. Viana, S. Zyngier, & G. Barnbrook, (eds.) *Perspectives on Corpus Linguistics* (pp.99-113). Amsterdam: John Benjamins Publishing Company
- Labov, W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia, PA: University of Pennsylvania Press
- Leech, G. & Svartvik , J. (2013). *Communicative Grammar of English*. London: Routledge.
- Leech, G. (2006). Principles and applications of Corpus Linguistics. In V. Viana, S. Zyngier, & G. Barnbrook, (eds.) *Perspectives on Corpus Linguistics* (pp.155-170). Amsterdam: John Benjamins Publishing Company
- McCarthy, M. (2004). *Touchstone from Corpus to Course Book*. Cambridge: Cambridge University Press
- Nelson, G. (2006). The Core and Periphery of World English: A corpus-based exploration. *World Englishes*. 16, 115-129
- Sinclair, J. (ed.) (1991). *Collins Cobuild English Grammar*. London: Collins CoBUILD
- The British National Corpus*. Retrieved February 3, 2015 from <http://corpus.byu.edu/bnc/>
- Thomson, A. & Martinet, A. (1986). *A Practical English Grammar*. (Fourth edition) London: Oxford University Press
- Viana, V., Zyngier, S. & Barnbrook, G. (eds.) (2006). *Perspectives on Corpus Linguistics*. Amsterdam: John Benjamins Publishing Company