A Corpus-based Cognitive Linguistic Analysis of Pre-existing Knowledge of Scientific Terminology: The Case of English *Energy* and Arabic *طَاقَة* (ṭāqa)

Hicham Lahlou  
Department of English Language Studies,  
School of Humanities  
Universiti Sains Malaysia, Penang, Malaysia

Abstract
The present paper aims to broaden the current understanding of students’ misconception of scientific terminology by identifying the gaps between Arabic and English scientific terminologies and between everyday language and scientific language. The paper compares the polysemy, prototypes, and motivating factors of English energy with those of Arabic *طَاقَة* (ṭāqa), with more focus on students’ prior knowledge. The study employs Lakoff’s (1987) idealized cognitive models and Rosch’s (1975) prototype theory to reveal the radial members of both categories, i.e., energy and *طَاقَة* (ṭāqa), and to explain the kinds of cognitive mechanisms that motivate the extension as well as understanding of the meanings of these terms. To this end, the study uses several English and Arabic dictionaries, lexical databases and corpora. This is to explore all the meanings, prototypes and motivating factors of the terms under investigation. The results show that the terms *energy* and *طَاقَة* (ṭāqa) overlap in prototypical meanings and motivating factors but differ in less prototypical and peripheral meanings. English and Arabic learners may then face similar issues in learning scientific concepts due to the difference between their pre-existing knowledge and scientific language.

Keywords: conceptual metaphor, conceptualization, energy, polysemy, prototype

DOI: http://dx.doi.org/10.24093/awejtls/vol4no1.1
Introduction

Among the multiple factors that can impact on the process of learning science, language has proven to be a significant aspect in understanding scientific concepts and ideas. Students usually make use of their own knowledge about scientific concepts taught in science classes, especially when such concepts are denoted by words used in everyday language. This prior knowledge is mainly based on learners’ cultural, social and bodily experience. Given the differences between scientific terminology and everyday speech, this knowledge conflicts with scientific concepts. These differences result in learners’ confusion and difficulties in understanding scientific concepts like NATURE, OBSERVATION, FORCE, ENERGY, SPEED, and ACCELERATION. Research into modern scientific terminology and its translation into other languages showed that learners might conceptualize scientific terms differently across cultures and languages (e.g., Kawasaki, 1996; Aikenhead & Ogawa, 2007; Lahlou & Hajar, 2013; 2016; Lahlou, 2018).

To see clearly the impact of this issue, let us consider an example from the aforementioned scientific concepts. Research on the term observation found that translating this term to other languages resulted in losing some of its basic concepts. For instance, English observation has inaccurately been rendered into Japanese as kansatsu (Kawasaki, 1999). The term observation entails that the relationship between the observer and the observed is objective as they are separate. The word kansatsu, in contrast, does not entail the isolation between the observer and the subject of observation. In another study by Aikenhead (2001) on the same Western scientific concept to observe, translated into Cree, an aboriginal language in Canada, as wapahtam, comparable distortion of the term’s underlying meaning was identified. First, to observe suggests all the human senses, whereas wapahtam refers to vision only. Second, unlike the Western scientific procedure to observe, wapahtam connotes a biased relationship between the observer and the subject observed.

Even though learners’ misconception of scientific terminology received much attention in the past few decades, research focuses on only few scientific concepts like nature and observation, especially in non-Western countries. Furthermore, studies on different scientific terms in Arabic are still lacking. Previous Western research, in contrast, was carried out on more scientific terms. For example, few studies investigated English students’ misconception of the term energy and found that students use their pre-existing knowledge of this concept in the science classroom and so fail to learn its primary scientific concept. According to these studies, students’ conceptualization of energy as fuel, substance, fluid, an object that people possess, etc. constitutes a misconception of the term in science contexts (e.g., Watts, 1983; Gilbert & Pope, 1986; Trumper, 1993).

The present paper further studies the linguistic effect on conceptualizing scientific terminology by examining the above-mentioned scientific concept in English and Arabic. Lakoff’s (1987) idealized cognitive models (ICMs) were utilized to obtain an exhaustive interpretation of the polysemy of English energy and Arabic طاقة (ṭāqa) and prototype effects caused by the projection of the ICM ENERGY. ICMs were employed because understanding word meanings is linked to an organized background of experience, beliefs, and so on, which forms a conceptual

Abstract

This study aims to analyze the influence of language on the conceptualization of scientific terminology, using Lakoff’s idealized cognitive models (ICMs) to obtain an exhaustive interpretation of the polysemy of English energy and Arabic طاقة (ṭāqa) and prototype effects caused by the projection of the ICM ENERGY. ICMs were employed because understanding word meanings is linked to an organized background of experience, beliefs, and so on, which forms a conceptual

Arab World English Journal for Translation & Literary Studies
ISSN: 2550-1542 | www.awej-tls.org
prerequisite to sense understanding (Fillmore & Atkins, 1992). Thus, the current study aims to compare the polysemy of energy and that of طاقة (ṭāqa). It also aims to find the commonalities and differences between the prototypes as well as cognitive mechanisms that motivate the conceptualization of energy and those that motivate the conceptualization of طاقة (ṭāqa).

Theoretical Framework
Three main theories were used in the present study. First, Lakoff’s (1987) ICMs were employed to elucidate the conceptual structure of English energy and Arabic طاقة (ṭāqa), including their by-products: category structures and prototype effects. Second, Rosch’s (1975) prototype theory is essential to illustrate the internal structure of a concept, that is, the prototypical and peripheral meanings of the category. In cognitive linguistics, polysemy, which the current study examines to identify the differences between the conceptual structure of English energy and its Arabic counterpart, is viewed as the result of the expansion of ICMs to create radial categories.

In cognitive linguistics, the background knowledge or encyclopedic knowledge against which concepts are understood is pivotal. Humans comprehend senses of words with recourse to their physical and social experience. For instance, the word bachelor signifies an unmarried adult male, but it may have several stereotypical implications concerning bachelor pads, dirty laundry, and so forth (Evans & Green, 2006). In cognitive linguistics, numerous methods of analysis incorporate the background knowledge, that is, Fillmore’s (1975) frames, Langacker’s (1987) cognitive domains, or Lakoff’s ICMs (1987). All these theories consider a language as a communication system which mirrors the world as human beings construe it (Cienki, 2007).

Ruiz de Mendoza (1999) defines an ICM as a conventional conceptual representation of the way human beings understand reality. It is a model since it attempts to be like reality. An ICM is how humans interpret it; it does not objectively exist in nature. It is cognitive because humans create it. It is idealized since it stems from many common experiences (Ruiz de Mendoza, 1999, as cited in Ibarretxe-Antuñano, 2004). The term Tuesday, for example, can only be defined through an ICM, which involves the natural cycle that is characterized by the movement of the sun and a seven-day calendric cycle. Days are linearly structured parts of the whole (week). Thus, Tuesday is the third part (day) of the whole (week). This model of a week is idealized; in other words, a seven-day week does not exist objectively in nature, but humans form it (Lakoff, 1987).

Prototype effects are the result of ICMs, which enable a speaker of a language to determine whether a category member is prototypical. The previously mentioned concept bachelor, for instance, can be understood relative to an ICM characterized by a monogamous community, youth, and adulthood. This ICM does not say anything about a priest. It is oversimplified because it “does not fit the world precisely” (Lakoff, 1987, p. 70). Within this ICM, some members of the concept of bachelor are exemplars. However, in contrast to Catholicism ICM within which a priest cannot marry, the marriage ICM entails a bachelor’s ability to get married. The dissimilarity between these ICMs then results in prototype effects (Evans & Green, 2006).
A metonymic model constitutes the most critical ICM that produces typicality. Lakoff (1987) describes metonymy as a case where a member of a category is employed to understand the category as a whole. He proposes many kinds of metonymic models that cause prototype effects, notably social stereotypes, typical examples, and salient (memorable) examples.

Social stereotypes are situations in which a member of a category is socially recognized to stand for the category as a whole (Lakoff, 1987). By way of illustration, a working mother is not a mother who works, but a mother who does not stay at home all day to nurture her children. The category WORKING MOTHER here is understood relative to stereotypical HOUSEWIFE-MOTHER. This stereotype metonymically stands for the category MOTHER as a whole and functions as a cognitive reference against which the other category members of MOTHER are defined. Consequently, WORKING MOTHER is a peripheral member of the category MOTHER.

Prototypical examples are the most common among the members of a category. In some cultures, for instance, APPLE and ORANGE are typical subcategories of FRUIT (Lakoff, 1987). All the other subcategories of FRUIT are defined in connection with the typical members. Salient examples are the frequent memorable subcategories that provide mental access to the whole category. For instance, California earthquakes stand for natural disasters. The present paper employed occurrence frequency due to its practicalities in research. In other words, the frequency of concurrence aids in identifying the typical members of a category (Geeraerts, 2006; Rosch, 1975).

Methodology
The current study compares the semantics and prototypes of energy and طاقة (ṭāqa), employing a set of selected English and Arabic dictionaries, lexical databases and two corpora: the British National Corpus (BNC) and the ArabiCorpus (Arabic Corpus Search Tool). Monolingual dictionaries and lexical databases (i.e., English-English dictionaries and Arabic-Arabic dictionaries) were used in the semantic analysis of energy and طاقة (ṭāqa). Table 1 shows the selected reference books and lexical databases used in the examination of energy and طاقة (ṭāqa).

<table>
<thead>
<tr>
<th>English-English</th>
<th>Arabic-Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Concise Oxford Dictionary</td>
<td>لسان العرب (Liṣān l-ʿ Arab)</td>
</tr>
<tr>
<td>Oxford English Reference Dictionary</td>
<td>القاموس المحيط (Al-Qāmūs l-Muḥīt)</td>
</tr>
<tr>
<td>Merriam-Webster Online Dictionary</td>
<td>المعجم الوسيط (Al-Muʿjam l-Wasīt)</td>
</tr>
<tr>
<td>Longman Dictionary of the English Language</td>
<td>معجم اللغة العربية المعاصرة (Muʿjam l-luġa l-ʿArabīyya l-muʿāṣira)</td>
</tr>
<tr>
<td>WordNet</td>
<td>المعجم الفني (Al-Muʿjam l-Ǧanī)</td>
</tr>
</tbody>
</table>
The Arabic Corpus (173,600,000 words) comprises newspapers, pre-modern text, modern literature and non-fiction. The BNC (100,000,000) consists of spoken, fiction, magazines, newspapers, academic texts, and so on. Both the BNC and the Arabic Corpus are beyond the minimal size of a standard corpus as they exceed ten million words (O’Keeffe et al., 2007; O’Keeffe & McCarthy, 2010). Furthermore, the two corpora are mostly equal in terms of language diversity since they both include a vast range of types of data, notably newspapers, fiction, non-fiction, spoken and written varieties of language.

The Arabic and English dictionaries and corpora were surveyed to explore the meanings of energy and طاقة (ṭāqa). The corpora also helped in deriving examples of these terms from the two comparable corpora and identifying the most frequent collocates of the terms concerned, which help in establishing the prototypes motivating their conceptualization. To this end, the collocates of energy and طاقة (ṭāqa) were extracted from the Arabic Corpus and the BNC respectively. The study focused on collocates in both the right and left co-texts. A span of up to four words on either side of the node was adopted. Sinclair, Jones and Daley assert, “beyond four words from the node there were no statistical indications of the attractive power of the node” (cited in Sinclair, 1991, p. 106).

Results and Discussion
The data on the terms energy and طاقة (ṭāqa) compiled from English and Arabic dictionaries and corpora were analysed. The results help identify the extent of similarity and difference between these words in terms of polysemy and prototypicality. This, in turn, helps delineate the way these terms are conceptualized.

The terms energy and طاقة (ṭāqa) are current words in everyday speech and formal language, and so their polysemy is complex. As mentioned before, dictionaries and lexical databases were consulted to examine the polysemy of these terms. To explore the nature of their polysemy, let us consider the following definitions.

Table 2. The Polysemy and Examples of energy and طاقة (ṭāqa)

<table>
<thead>
<tr>
<th>Polysemy</th>
<th>English</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) ‘ability or capacity for doing work’ / ‘the physical and mental effort used to do something, dynamism, drive’</td>
<td>...they [retired people] will never be able to acquire the creative energy necessary to find a new job.</td>
<td>يحاول جاهدا باذل كل طاقته في تنفيذ هذا الشيء. yuḥāwilu jāhidan bāḏila kulla ṭāqatihi fi tanfīḏi haḍā aš-šay’ He tries hard, putting all his energy into the fulfilment of this.</td>
</tr>
<tr>
<td>(2) ‘window’</td>
<td>NIL</td>
<td>كانت الصحف هي النافذة والطاقة والوسيلة إلى معرفتها والقاء الضوء عليها [المشاكل]. kānati aṣ-ṣuḥufu hiya an-nāfiḏa wa aṭ-ṭāqa wa l-wasīla ’ilā ma’rifatihā wa ʾilqāʾi aḍḍaw’i ʿalayhā [al-mašākīl]</td>
</tr>
</tbody>
</table>
Newspapers were the window and means whereby they [issues] were identified and clarified.

| (3) 'usable power'/‘a source of power’ | They then build the plants required to generate energy. | مصارف دولية تقدم 828 مليون دولار لتمويل مشروع مغربي لتوتّل الطاقة. maṣārifa dawliyatan tuqaddimu 828 malyūn dūlār li-tamwīli mashrū’ in maġribiyiyin li-tawfīdi aṭ-ṭāqa International banks provide $ 828 million to finance the Moroccan project to generate energy. |
| (4) ‘capacity of production’ | NIL. | وتبلغ الطاقة الإنتاجية للمصنع 500 ألف طن متروي سنوياً. wa tabluġu aṭ-ṭāqatu lʾintājiyyatu li-ilmāsna’ 500 ʾalfa ʾṭan mitrī sanawiyyan The production capacity of the factory is 500 thousand metric tons per year. |
| (5) ‘bunch, bundle, bouquet’ | NIL. | قدم قتى وفاة من ابناء الشهداء طاقتين من الزهر للسيد الرئيس. qaddama fiṭī wa fiṭāt min abnāʾi ššuhadāʾ ṭāqatayni mina z-zahr li-s-sayyid r-raʾīʾs …A martyr’s son and a martyr’s daughter presented a bouquet of flowers to Mr. President. |

Considering the various meanings of the terms *energy* and طاقة (ṭāqa) shown above, it is remarkable that they share some denotations. Both mean ‘ability’ and ‘usable power, or source of power’. The English and Arabic denotations of the category ENERGY, that is, ‘ability’ in (1) in Table 2 can be understood in conjunction with the connotations of or encyclopedic knowledge about ENERGY. Within the ICM of ENERGY, *energy* and طاقة (ṭāqa) can be understood in connection with such elements as ambition, action, physical strength, mental strength, and so forth. The denotations of ‘usable power’ in (3) in Table 2 can be comprehended in connection with elements like electricity, powering machines, dams, power, light, and heat.

However, the words *energy* and طاقة (ṭāqa) differ in some other meanings. Unlike *energy*, طاقة (ṭāqa) can be used with the meaning of ‘a window’, ‘a bunch’, and ‘capacity of production (preceded by إنتاجية (ʾintājiyya) (productive))’. In Example (2) in Table 2, طاقة (ṭāqa) is projected to include the meaning of ‘a window’, and thus connotations like building, frame, glass and light surface in the conceptualization of طاقة (ṭāqa) as a window. In example (4) in Table 2, طاقة (ṭāqa) is extended from *energy* (source domain) to *capacity* (target domain) with the addition of the adjective إنتاجية (ʾintājiyya) (productive). Within BUSINESS ICM, طاقة (ṭāqa) can be comprehended in terms of features like enterprise, resources, workforce, products, and economy. In example (5) in Table 2, طاقة (ṭāqa) is projected to include the meaning of ‘bouquet or bunch’. This subcategory of طاقة (ṭāqa) is associated with many elements such as similar things contained, collection, present, and flowers.
As shown in Table 3, the data compiled from the BNC and ArabiCorpus show that the most frequent collocate of *energy* is *efficiency* and the most frequent collocate of طاقة (tāqa) is نوروية (nawawiya) (atomic).

Table 3. The Most Prototypical Collocates of *energy* and طاقة (tāqa)

<table>
<thead>
<tr>
<th>Collocates in Freq. Order</th>
<th>The British National Corpus</th>
<th>The ArabiCorpus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word</td>
<td>Collocate</td>
</tr>
<tr>
<td>1st Collocate</td>
<td></td>
<td>efficiency</td>
</tr>
<tr>
<td>2nd Collocate</td>
<td></td>
<td>atomic</td>
</tr>
<tr>
<td>3rd Collocate</td>
<td></td>
<td>conservation</td>
</tr>
<tr>
<td>4th Collocate</td>
<td></td>
<td>department</td>
</tr>
<tr>
<td>5th Collocate</td>
<td></td>
<td>nuclear</td>
</tr>
</tbody>
</table>

Thus, *energy* is commonly conceptualized as the efficient use of resources, evoking several features like energy resources, economy, sustainability, and so on. A less frequent collocate of *energy* is atomic, virtually interchangeable with nuclear. In this sense, *energy* can be associated with a source of energy, electricity, agency, weapons, and so forth.

In contrast, the most frequent collocate of طاقة (tāqa) is نوروية (nawawiya) (atomic), which is exchangeable with نوروية (nawawiya) (nuclear). Despite the dissimilarities between atomic and nuclear, mainly in physics, they are similar when they modify energy. This indicates that the Arabic term طاقة (tāqa) is frequently associated with forms or source of energy and with weapons. Conceptualising طاقة (tāqa) as a source of energy is supported by the other common collocates, notably كهربائية (kahrabā’iya) (electrical) and شمسية (šamsiya) (solar). A less frequent collocate of طاقة (tāqa) is إنتاجية (’intājiyya) (productive), which shows that the term طاقة (tāqa) can be comprehended based on (the capacity of) production.

The results, thus, show that there are many cases where the English term *energy* and the Arabic term طاقة (tāqa) are parallel, especially in denotations, including ‘ability’ and ‘useable power’. However, only طاقة (tāqa) has denotations like ‘a window’, ‘a bunch of’, and ‘production capacity’. This shows that the polysemy of Arabic طاقة (tāqa) has a more complex conceptual structure than that of English *energy*. 
In the first commonality between energy and طاقة (ṭāqa), ENERGY entails force or rather a forceful exertion, which relates to physical as well as mental abilities:

6. Tea and herbs and lemon to sharpen it, not too sweet but sweet enough to put some energy into her.

7. أريد أن أغفو قليلا، فلم تبق في جسدي طاقة…
I want to sleep a little, no energy is left in my body…

8. …she develops freely physical and mental energy and will not be physically exhausted and mentally debased.

9. …She develops freely physical and mental energy and will not be physically exhausted and mentally debased.

In these examples, energy and طاقة (ṭāqa) denote physical as well as mental vigour. In examples 7, 8 and 9, the exerters I, she and a ten-year-old boy exert energy respectively while in example 6 tea, herbs, and lemon give her body a burst of energy.

The second sense of energy and طاقة (ṭāqa), i.e., ‘usable power’ or ‘source of power’, can also be understood in terms of substance. That is to say, the source domain FUEL is mapped onto the target domain ENERGY. In this regard, energy is the outcome of fuel production.

10. That’s, now that fuel is producing energy.

11. …lack of linking scientific research to the development of local industries, such as medicines and fuel (energy) alternatives.

This kind of understanding not only entails electricity and petrol, but also food.

12. …so we can measure the energy that food provides in calories.

13. ...Food is the energy that enables a human to move.
The conceptualization of طاقة (ṭāqa) as fuel and source of power is supported by the collocate frequency data discussed earlier. English energy is also evidently conceptualized as fuel given that its second most frequent collocate is atomic. However, while طاقة (ṭāqa) is closely associated with the generation of power, energy is strongly associated with the efficient utilization of resources, that is, less energy consumption to offer similar service. Despite the difference in these two processes, they both concern power or rather source of power. Thus, the prototype that motivates the meaning of energy and طاقة (ṭāqa) is a source of power.

Though the most prototypical entity that may be the primary motivating factor for the conceptualizations of both energy and طاقة (ṭāqa) is a source of power, there are some differences at the level of the periphery. The word طاقة (ṭāqa), in contrast to energy, can mean ‘production capacity’, which is less prototypical based on the data compiled from the ArabiCorpus. The corpus data on energy and طاقة (ṭāqa) also indicates that the dissimilarities in the meanings ‘a window’ and ‘a bunch’ are no more significant as these senses have become rare. In 44,232 occurrences of طاقة (ṭāqa), the meaning of طاقة (ṭāqa) as ‘a window’ was found in one concordance line only. In the same vein, the meaning of طاقة (ṭāqa) as ‘bouquet’ was found in one concordance line only.

Conclusion

The current study showed that there are commonalities and differences between the term energy and the term طاقة (ṭāqa) in terms of polysemy and prototypicality. The word طاقة (ṭāqa) is more polysemous than the word energy; however, these terms have the same prototypical meanings but differ in peripheral senses, most of which are rarely used in Modern Arabic. The terms are then virtually identical except for some minor differences in less prototypical and non-prototypical senses. Thus, these differences are deemed insufficient to impact on Arab learners’ understanding of the concept of ENERGY.

Arab and English learners are then presumed to face the same difficulties in learning the concept concerned. More precisely, these learners may not understand the exact meaning of the concept ENERGY in science classes because of the pre-existing knowledge they have about this concept. The intended meaning connected with the scientific cognitive model does not form part of the existing knowledge and is mostly unrelated to the prior knowledge in a student’s mind. Conceptualizing energy as human property, substance, source of power, or fuel is not congruent with PHYSICS ICM. Research on students’ misconception of the term energy demonstrated that these types of understanding might hinder students’ learning of this concept in a scientific context (e.g., Watts, 1983; Gilbert & Pope, 1986; Trumper, 1993).

Humans may thus have conflicting ways of understanding. This is because there are folk and expert theories of science, which include idealized cognitive models in correspondence with a language (Lakoff, 1987). The fact that idealized cognitive models do not match one another is ordinary, but the challenge may be in applying a cognitive model in the relevant context. Therefore, to facilitate learning scientific concepts, the gaps between everyday language and science language should be emphasized. This will help students to avoid understanding science terminology based on everyday conceptions, to be aware of the gaps between folk and expert theories of science, and to be able to choose the consistent ICM for a specific range of contexts.
A Corpus-based Cognitive Linguistic Analysis of Pre-existing Knowledge

About the Author
Hicham Lahlou (PhD) is a faculty member at the School of Humanities, Universiti Sains Malaysia. He has more than twenty-two years of experience in teaching high school and university English. His research interests lie within English linguistics, cognitive linguistics, corpus linguistics and education. https://orcid.org/0000-0003-2796-9877

References


---

In this paper, words like energy are in italics, senses such as ‘ability or capacity for doing work’ are in quotation marks, transliterated as well as translated words, like طاقة (ṭāqa) (energy) are given in brackets, and concepts such as ENERGY are written in small caps.

---

The Arabic words, in this paper, are transliterated in accordance with the recommended transliteration method used in academic studies of linguistics (Isaksson, 2013).