Linguistica ONLINE
Issue Twenty Three
ISSN 1801-5336

Miscellanea

XII

May 2020 • Department of Linguistics and Baltic Languages, Masaryk University

Linguistica ONLINE

ISSN 1801-5336
electronic journal of the Department of Linguistics and Baltic Languages, Masaryk University, Czech Republic

home: **http://www.phil.muni.cz/linguistica/**
email: **linguistica@phil.muni.cz**

MISCELLANEA XII

http://www.phil.muni.cz/linguistica/art/issues/issue-023.pdf
published: May 7, 2020

CONTENTS

Issue Twenty Three

http://www.phil.muni.cz/linguistica/art/issues/issue-023.pdf

Linguistica ONLINE. Issue 23. Published: May 7, 2020
http://www.phil.muni.cz/linguistica/art/kfouri/kfo-001.pdf
ISSN 1801-5336

ANALYZING SPEECH ACTS IN LEBANESE ARABIC, FRENCH AND ENGLISH BEER ADVERTISEMENTS: A CROSS CULTURAL STUDY OF SPEECH ACTS[*]

Carla Kfouri (Beirut Arab University, kfoury.carla011@gmail.com)

Abstract: Speech act analysis won the interest of linguists, philosophers, and sociolinguists over time after the emergence of Austin and Searle's speech act theories in the sixties. Language was no longer investigated in isolation from the context it belonged too. Hence, analyzing speech acts within advertisements gave more and more insight into how speech acts are used across different cultures. Though the content of advertisements is carefully selected by promoters to attract a specific audience, the choice of speech acts usually takes into consideration the cultural framework that defines a particular community. For that reason, the present study attempts to analyze linguistically and cross culturally speech acts that are implemented in Arabic, English, and French beer written advertisements. This is achieved through Searle's (1969) model of speech act theory and qualitative content analysis. The purpose is to provide additional insight into how speech acts are implemented in different cultures within beer advertisements and subsequently compare the existent cultural elements within the latter. Thirty written beer advertisements divided equally in Arabic, English and French language were selected from various websites. After classification of speech acts and identification of themes through content analysis, dominant themes in advertisements were classified based upon Hofstede's (1980) and Pollay's (1983) cultural value system, the findings reveal that English ads deploy directives more than assertives compared to Lebanese Arabic and French beer ads where assertives seem to be prioritized. Thematically, English ads focus on individualism and self-satisfaction whereas Lebanese ads focus more on collectivism and patriotism French beer ads focus on origin and country of product which could be related to patriotism as well. This indicates that though speech acts are universal, their usage relies heavily on the cultural values of the country.

Key words: Advertisement, culture, speech acts, themes

1. Introduction

Language is not just a tool of communication but an effective means of bringing thoughts together, improving society and reflecting changes upon communities. It is actually part of the wider socio-cultural context where the aim is "to look into language from the outside and specifically, to interpret linguistic processes from the standpoint of the social order"

[*] Previously unpublished. [Editor's note]

(Halliday 1978, p. 3). In each context, language serves multiple functions like raising awareness, advocating concepts and promoting items. For instance, the language of advertisements shares that last purpose promoting items and objects to increase the selling process. It is widely known that advertising is in fact a tool to promote products that companies manufacture with the purpose of selling within a competitive industrial environment. The main purpose of advertisements is to attempt to persuade potential customers to purchase or to consume more of a particular brand of products or services (Widyatama, 2007, p. 141). In order to fulfill its targets, there exists a strategy that most commercial advertisements adopt which lies in the message itself. The message has a crucial role in the advertising process. It is the core of advertisement (Kotler, 2000, p. 13).

If we examine messages in a commercial advertisement, we can easily notice that they are usually conveyed through two ways. The first one consists of the visual part where there is an image that depicts the content of the advertisement. This image reflects the product it is promoting to catch the eye of the consumer and increase the selling process. On the other hand, commercial advertisements also use written messages to address their audience or potential buyers. The written messages often comprise speech acts to call for action beside other purposes. However, this approach might prove risky because if the message is not accessible and comprehensible, loss of interest and vagueness might arise. This is what compels advertisers to carefully choose the content of their advertisements by referring to poetic and linguistic devices like speech acts to enhance the idea they want to present to the consumers. If the message is not transmitted well, the communication process will be disturbed (Kotler, 2000, p. 3). Therefore, to guarantee their success, advertisers have to study carefully the needs as well as the cultural background of the customers.

1.1 Background

Advertisements are in general defined as cultural mirrors, which reflect the culture in question, but also most familiar and stereotypical ideals, norms and values. In her book *Advertising as Communication*, Gillian Dyer (1982) points out that most advertisements "present what appears to be the only ideal and the desirable way of living" (Dyer, p. 11). Advertising and culture are closely connected to the extent that advertisements can never be analyzed without reference to the cultural background of the audience that those advertisements target. Within that scope, Marieke de Mooij (2005) states that we all have universal needs, but the way we satisfy them differs among cultures" (de Mooij, p. 37). Advertising may vary across cultures not in brand specifically but in the usage of speech acts since generally speech acts reflect the cultural make up of society. Though previous studies have tackled cross cultural analysis of advertisements in different countries, there is still need of further analysis of advertisements in Arabic language as compared to English and French advertisements to investigate whether speech acts are universal or not. This issue needs further investigation since cultural differences may impose differences in the use of speech acts. Therefore, commercial ads have to take into account the cultural elements that exist in different countries.

1.2 Purpose

The purpose of this study is to compare and contrast the types of speech acts implemented in Arabic, English and French beer ads and subsequently draw a cross cultural comparison between them. This is achieved based on Searle's taxonomy of speech act classification and content analysis. The implementation of qualitative analysis of speech acts and content analysis of English, Arabic and French beer ads leads to a clearer picture regarding how advertisements are realized in some eastern and western countries. Furthermore, the study focuses on the cultural impact embodied in those advertisements and how the use of speech acts across cultures reveals different cultural elements. Taking this into consideration might help advertisers realize the needs of each society; thus improving the quality of commercial ads to increase sales in a particular country. This might also be important in avoiding cultural clash through transferring values that some people would consider strange in their community.

1.3 Limitations

The limitations of the study lie in the small sample offered from different cultures since the article is exploratory raising awareness about cultural issues through the analysis of speech acts in a number of ads. There are other limitations to be addressed including the lack of interviews with advertisers and recipients of the advertisements. This could be implemented as well for future research.

1.4 Research questions

1. What are the types of speech acts that exist in Arabic, English, and French beer written ads?
2. What are the similarities and differences that exist in Arabic, English, and French beer ads?
3. What are the cultural elements underlying Arabic, English, and French ads?

2. Literature Review

2.1 Speech act theory

Messages in commercial advertisements usually comprise speech acts since speech acts according to Searle (1969) speech act theories extend to being used in different pragmatic situations to express states and actions (Muller, 2016, pp. 3–7). This implies that speech acts work the best within their social context that grants them their effectiveness and functionality. Though the use of speech acts might seem artificially implemented within advertisements, nevertheless, these advertisements target a specific audience that is part of

a society and part of a culture. Hence, promoters are the main source behind these advertisements and are indirectly socially involved with their target audience. This is what Searle attempted to convey through his elaboration of speech act theory. Speech acts might extend to different types of communication including advertisements and derive their meaning from the cultural make up of a particular society.

Searle (1969) suggested replacing Austin's locution / illocution distinction with utterance acts in which the speaker utters words, propositional acts in which the speaker performs illocutionary acts, which have a particular force. He set some more detailed rules concerning felicity condition for each illocutionary acts. According to him, several conditions have to be fulfilled for a sentence to be felicitous. These rules mostly concern the psychological beliefs of the speaker or hearer and each one of them has to be fulfilled in order to create a felicitous act. These rules are propositional content, preparatory condition, sincerity condition, and essential condition as explained in the following:

1. Propositional content: The propositional content condition explains the illocutionary forces and specifies the acceptable conditions regarding propositional content. In other words, it is the proposed condition of the speaker or hearer.
2. Preparatory condition: In an attempt to conduct a felicitous illocutionary act the speaker has to have certain beliefs about the speaker's act and conditions, and is also required to have the power of authority or persuasion over the hearer.
3. Sincerity condition: In performing a felicitous act, the performer must have a certain psychological attitude concerning the propositional content of the utterance. For example, when a person is making a promise, he/she must have an intention of keeping it.
4. Essential condition: Essential condition of an utterance has to do with its intention to get the hearer to perform the intended act.

According to Searle, two different speech acts may have the same propositional content and different illocutionary forces. For example, the propositional content of "I will go to bed early" can be to make a promise or a statement. This implies that each kind of speech act obtains its fulfillment depending both on the propositional content and the illocutionary force. If I make a promise, it is fulfilled if I keep it; but if I make a statement, it is fulfilled if it is true. To perform a speech act is thus to generate a propositional content linked to an illocutionary force. But to generate an illocutionary force, there must be felicity conditions (pp. 35–38).

Searle's speech acts comprise assertives that commit the speaker to something such as suggesting, putting forward, swearing, boasting, and concluding. Directives make the addressee perform something like asking, ordering, requesting, inviting, advising, and begging. Commissives commit the speaker to doing something in the future (promising, planning, vowing, betting, opposing). Expressives express how the speaker feels about a situation, such as thanking, apologizing, welcoming, deploring, and declarations that change the state of the world in an immediate way. Since language might lead to intentional behavior, it should be treated like a form of action. Being part of pragmatics, speech acts analysis reveals the intentionality of messages conveyed within a context (Searle, p. 39).

Searle (1969) further elaborates the notion of context relating speech acts to their cultural relevance to chess. Hence, he considers that speech acts are universal just like chess game. The rules of speech acts are the same but what differs is the language they use and the context those speech acts are carried out. He states: "First, imagine that chess is played in different countries according to different conventions. Imagine, e.g., that in one country the king is represented by a big piece, in another the king is smaller than the rook. In one country the game is played on a board as we do it, in another the board is represented entirely by a sequence of numbers, one of which is assigned to any piece that moves to that number. Of these different countries, we could say that they play the same game of chess according to different conventional forms" (Searle, 1969, p. 39). Context really counted for Searle. Therefore, the context is what gives meaning to the sentences and not the opposite (Schiffrin, pp. 6–16). This implies that speech acts have the same classification but their cultural usage might differ.

Applying Searle's theory in its notion of contextuality of speech acts, it becomes evident that the designers of beer advertisements are aware of the culture that they are addressing and attempt to simulate the occurrence of speech acts within their natural context in a different scenario that of a beer advertisement. The expected response is to leave influence on the customer but mostly call for action.

Another interesting theory is that of Blum-Kulka et al. (1989) stating that some universal pragmatic features of speech acts are not just universal, but also subject to culture-specific variations. Blum-Kulka et al. (1989) say if claims for the pragmatic universality are to approximate any type of validity, they should be based on the empirical investigation of many more and diverse languages" (p. 8). Hence, they suggest investigating speech acts in a variety of languages and cultures other than English. This is what Gas & Neu suggest doing since they consider that speech acts are realized from culture to culture in different ways, and these differences may result in communication difficulties that range from the humorous to the serious (p. 150). Similarly, Gass & Selinker (2008) claim that all languages comprise speech acts and presumably speech acts themselves are universal, yet the form used in specific speech acts varies from one culture to another. This explains the presence of thousands of studies done around speech acts in different languages and different contexts as reflected as well in the present article.

2.2 Previous Studies on cross cultural advertisement

Many studies have been conducted on speech acts in media contexts though fewer studies have investigated speech acts across cultures. The following studies provide an insight into what sociolinguists have achieved in the field of speech analysis across cultures. Chiluwa (2006) attempts, in a study based upon Searle's theory of speech act, to interpret the language of campaigns in terms of persuasive actions common to most marketing communications. The data comprised 20 adverts in the print media and outdoor advertising namely news-papers, magazines, billboards, posters and flyers between 2000 and 2006. Analyses show that the directive act and the representational act were more frequently demonstrated in the advert. The two acts involved requesting, appealing, claiming, reporting, asserting and suggesting. These appeared to be the best persuasive strategy in the adverts as they

particularly appealed to the consumer's sense of ego, sex, class and self-worth (pp. 29–38). Another study was led by Rahman (2008) who analyzed illocutionary acts in the movie *Monster in Law*. The study applied discourse analysis by using quantitative and qualitative descriptive method. His research was based on Austin's theory and the types of sentences and performative verbs in the illocutionary acts performed in the movie *Monster in law*. The objective of this research was finding out the performative verbs of the illocutionary acts that consist of verdictive, exertive, commissive, behabitive and expositive and types of sentences that consist of declarative, interrogative, exclamatory, and imperative sentences in the movie *Monster in Law*. The findings implied that there were four types of sentences such as declarative sentence, interrogative sentence, imperative sentence and exclamatory sentence. The performative verbs were verdictives, commissives, behabitives and expositives (pp. 6–20).

Cemalovic (2009) conducted a comparative analysis between Danish and British beer advertisements where the results showed a greater extent of similarities than differences. Although both cultures seem alike, they are dominated by different values and norms, which are reflected in Hofstede's and Hall's classification of the two cultures. In both cultures beer consumption is seen as a social phenomenon. It could be in some cases part of the national identity and consumed in company with others denoting relaxation. Coming to the differences, the Danish advertisement might depict beer consumption as being harmonic and reflective of feminine values while the British advertisement might be more focused on the high quality of the beer matching with masculine standards. This corresponds with Hofstede's classification of Denmark as a highly feminine culture, which values community and care for others. On the other hand, Great Britain could be more success-oriented due to its masculine values. Furthermore, the focus on the quality and the exclusivity of the beer could be related to the declining beer consumption, where Carlsberg is intending to reinforce a sophisticated and luxurious image of the beer (p. 10–31).

Naufalina (2017) examined speech acts used in *Bon Appetit* food advertisements by identifying the types of speech acts, and figuring out the speech act patterns in the advertisements. This research was conducted using a descriptive qualitative method. The data were collected from *Bon Appetit* published in March to May 2016 in the form of words, phrases, clauses, and sentences. After being collected, the data were analysed based on Searle's theory of speech acts. The results imply that statements serve as the highest type of locutionary acts which are aimed at providing information of the products, commissives serve as the highest type of illocutionary acts aimed at offering the benefits of the products, and getting the readers to expect something.

What can be concluded from all the studies that were conducted in speech act analysis within media and advertisements is that within the domain of advertisement, cultural elements affect the use of speech acts in different countries. Hence, in order to achieve their ultimate purpose, advertisements need to take into account the cultural values of a particular community.

Moreover, the choice of the latter studies in particular aim at pinpointing the social media context within which these speech acts occur ranging from advertisements to movies, newspapers and magazines. The occurrence of speech acts within these frames is not natural but devised by the promoters of these ads who attempt sometimes to imagine scenarios that reflect certain aspects of real life and expect similar responses. As has been shown from the

findings of those studies, cultural differences imply different types of speech acts to approach a particular audience. However, the functions of those speech acts might be somehow similar and more than often might imply a commitment, call for action or simply provoke desire.

2.3 Some aspects of Eastern versus Western culture

Different cultures around the world have different norms, behaviors, and actions, which in essence are different ways of communicating with one another according to their rules and principles. However, different actions and behaviors carry different meaning depending on the context and the culture. According to Hymes (1962), part of being culturally competent speaker is understanding the speech events recognized by that culture (pp. 99–138). By this idea, Hymes (1962) implies that speech events are governed by norms or rules for the use of speech within a speech community. The speech event might consist of a single speech act, or it may comprise multiple speech acts. Examples of speech events include jokes, conversations, and arguments among many others and usually take place within a speech situation that is social or cultural context.

Hall & Hall (1987) consider that context is the inseparable piece of information that is surrounding an event and helps in giving the event some meaning. Different cultures communicate with different context, and that is why cultures can be compared on a scale from low to high context (Hall & Hall, p. 6). They define high context communication as where most of the information exists in the person and there is very little explicit information in the transmitted message while low context communication has most of the information in the explicit and transmitted part of the message. Cultures that prefer to have close personal relationships and have extensive information networks among colleagues, family, and friends are high context, such as some Arab and Japanese people. Therefore, the members of these cultures do not need to have background information about any activities or events since they keep themselves informed of what is going on with others especially if the people involved are considered important. On the other hand, Hall and Hall (1987) consider that cultures that separate themselves and stay discrete about every aspect of their lives are low context cultures, such as some Americans and Germans. As a result, they find themselves in need of detailed background information when interacting with others (pp. 8–12). Whereas Hall & Hall focus on low and high context cultures which affect the way individuals of a particular culture interact with each other and with other cultures, Hofstede (2011) views culture differently. He defines culture as a collective programming of the mind distinguishing the members of one group or category of people from others. He also developed a model that consists of several dimensions that make up every culture. These dimensions are power distance, masculinity versus femininity, individualism versus collectivism, long versus short term orientation, indulgence versus restraint, and uncertainty avoidance (p. 5). Hofstede (2011) distinguishes five primary differences in cultural patterns and how they influence language.

The first one is power distance. Power distance index focuses on the degree of equality or inequality, between people in the countries. A high power distance ranking indicates that there exist inequalities of power and wealth within societies. These societies are supposed to follow a system that does not allow up citizens to rise up in class or power. A low power distance ranking implies that the society overlooks the differences between the citizen's

power and wealth. In these societies, everyone has equal opportunity. Because of the influences of feudalism culture, China is a high power distance ranking nation. America is a low power distance ranking nation because they believe that everyone is equal before God.

The second cultural pattern that I used some aspects of in my study is Hofstede's focus on individualism and collectivism since this reflects on the usage of speech acts by people of different cultures. In other words, if people belong to an individualistic culture, they might use different speech acts that reflect that individualism (pp. 10–20).

Hofstede (2011) states that individualism pertains to societies in which the ties between individuals are loose. Everyone is expected to look after himself or herself and his or her immediate family. Collectivism, on the other hand, pertains to societies in which people from birth onwards are integrated into strong cohesive in-groups, seeking protection and loyalty. In individualistic cultures, like America, people value their "I", and their independence and self- reliance, whereas in collective cultures, like China, people are more group-oriented and concepts such as "we" interdependence, duties, obligations, needs, and views of in-groups are cherished. Individualism and collectivism exist in all cultures. This permits members of individualistic cultures to acquire some collectivistic values as well. It is worth mentioning here that Hofstede's theory could be subject to criticism and inconclusive but I have implemented the general cultural dimensions of Hofstede in my analysis of Lebanese, English and French ads.

Masculinity-femininity is another cultural pattern. Masculinity means the culture in which the dominant values in a society are male oriented and is associated with behaviors such as ambition, sex rules, achievements and the acquisition of money. Femininity is a culture where man values equal responsibilities with a woman such as nurturing roles and believing in sexual equality. Therefore, a high masculine society is one where the man acquires superiority over a woman in all aspects of life whereas a low masculine society is one that values women and their rights. Since the 1960's, the women's liberation movements began to rise in western countries. Women called for the same rights as those granted to males. This is reflected in language itself. For instance, in the past, there was only the word "chairman" and "businessman" but nowadays the words "chairwoman" and "business-woman" came into use.

Within that scope, Hofstede (2015) claims that the collectivism side of the dimension measures the degree to which individuals in a society are united in groups. Furthermore, individuals in such societies that fall on the collective side are integrated strongly into groups that are often extended families that protect them, and in return the family expects unquestioned loyalty. Moreover, societies that fall on the collectivism side prefer tightly-knit social framework, and individuals' self-image is defined as "We". This implies that in societies that are collectivist, maintaining harmony is of high importance. On the other hand, individualism is a focus on rights above duties, a concern for oneself and immediate family, an emphasis on personal autonomy and self-fulfillment, and the basing of one's identity on one's personal accomplishments. Individuals are expected to look after themselves and their immediate families, and the reason is that ties between individuals are loose. In other words, individualist societies prefer loosely-knit social framework, and individuals' self-image in such societies is "I" (pp. 11–30).

The fourth cultural pattern is uncertainty and avoidance. This cultural pattern focuses on the level of tolerance for uncertainty and unpleasant situations. Some nations know how to

cope with uncertainty through different ways. For instance, America is a nation whose culture is characterized by low uncertainty avoidance. This means that Americans do not feel so threatened or anxious about uncertainty. Hence, they do not feel the urge to limit or control it. This can be reflected in their language just like in the following examples: "Risks are opportunities" and "conflict in organization is natural, nothing is to be afraid of".

The last cultural pattern that Hoefstede discussed was time orientation. In Chinese culture, people believe strongly in that the past should be the guide for making decisions just like in the proverb: "Consider the past, you'll know the present." America is a future-oriented nation where people emphasize the future rather than the present. This is of course a generalization as, in every society, exceptions might arise and exist.

Many researchers have applied Hofstede's dimensions to position their studies, and have mostly focused on comparing countries that are completely distant in terms of region, culture and economic status, such as Eastern countries and Western countries (Dahl, 2004). However, the model of cultural values of Hofstede did not prove sufficient for measuring all aspects of cultural elements. Shortly later, Pollay (1983) described advertising as a carrier of cultural values and proposed a methodology for measuring the cultural values manifest in advertising by synthesizing and examining the work of previous authors. As a result, Pollay identified 42 advertising appeals such as adventure, affiliation, casual, cheap, community, convenient, dear, distinctive, durable, effective, enjoyment, family, frail, freedom, healthy, humility, independence, magic, maturity, modern, modest, morality, nature, neat, nurturance, ornamental, plain, popular, productivity, relaxation, safety, security, sexuality, status, succorance, tamed, technological, traditional, untamed, vain, wisdom, youth. Three years later, he developed the distorted mirror metaphor implying that advertisement reflects only certain attitudes, behaviors and values… those that serve sellers' interests (Pollay & Gallagher, 1990, p. 360). Hence, Pollay & Gallagher identified 25 cultural values frequently depicted in North American ads that manifested high consistency over time and across media (p. 359).

In short, the above mentioned cultural frameworks contributed to the development of research into cultural values in advertisement. Among them, Hofstede's and Pollay's cultural framework models proved to be the most effective for value measurement in cross cultural advertisement.

3. Method

3.1 Participants

There are no participants in this study because the participants are replaced by the content of the Arabic, English and French beer advertisements.

3.2 Design

The study is a linguistic and cross cultural one utilizing a qualitative method of content analysis because it fits the purpose of the study in determining the types of speech acts used in beer advertisements plus their themes and the cultural similarities and differences

embedded in the messages. The study is cross cultural since it deals with cultural elements across various cultures in written messages of beer ads.

Searle's (1969) speech act theory is implemented in categorizing speech acts in the written messages of Arabic, English, and French commercial advertisements. Searle's taxonomy of speech acts permits the classification of types of speech acts into their respective functions while content analysis allows theme classification. From theme classification, cultural elements could be pinned down based on Hofstede's-Pollay's framework. Hence, the design relies on a combination of Hofstede's (1980) cultural dimensions and Pollay's (1983) concept of values.

3.3 Materials

The material consisted of thirty beer advertisements, divided equally between Arabic, English, and French.

3.4 Data Collection

The Arabic, English, and French advertisements were collected online from different websites like pinterest, google, and others. The selection done was based on speech acts and thematic content.

4. Data Analysis

Since twenty advertisements were in Lebanese Arabic and French, I had to translate them all into English. Keeping the written text in Arabic and French would have prevented me from accurately comparing the speech acts with their English counterparts.

The translation process required many readings of each written text since a literal translation would affect the meaning. This is generally considered as an important first step in the process of analysis. After the completion of the translation, I started classifying speech acts based on Searle's taxonomy of speech acts. This was applied on all Arabic, English, and French advertisement written messages. The second step consisted of applying qualitative content analysis to reveal the themes that are embedded within the campaigns. According to Zhang and Wildemuth (2009), qualitative content analysis usually uses individual themes as the unit for analysis, rather than linguistic units (e.g., word, sentence, or paragraph) most often used in quantitative content analysis. Themes can be expressed in a single word, a phrase, a sentence, a paragraph, or an entire document (p. 310). "Theme" could be described as the subjective meaning and cultural-contextual message of data. Codes with common points of reference, a high degree of transferability, and through which ideas can be united throughout the study phenomenon can be transformed into a theme. In other words, a theme is a red thread of underlying meanings, within which similar pieces of data can be tied together and within which the researcher may answer the question "why?" (Erlingsson &

Brysiewicz, 2013). Determining the themes might help in mapping the cultural elements that exist in advertisement messages.

The coding frame was based on Hofstede's (1980) and Pollay's (1983) framework of advertising cultural values. Each message was coded for one dominant theme. The themes were then grouped under one major theme and then classified according to Hofstede's and Pollay's cultural values paradigm.

5. Results

a. English advertisements

1. *Log off. Lime in. Find your beach.*
   Speech act: Directive.
   Theme: Seeking a break (Relaxation)
2. *Enjoy life with Miller high life*.
   Speech act: Directive.
   Theme: Individual pleasure.
3. *Why is Heineken America's number one imported beer. Taste*.
   Speech act: Directive.
   Theme: Sensory pleasure.
4. *Oceans apart from the ordinary*. *Heineken tastes tremendous-no wonder it is number one.*
   Speech act: Assertive.
   Theme: High quality.
5. *Heineken tastes tremendous*.
   Speech act: Assertive
   Theme: High quality.
6. *She found she married two men*. *She found she has it all.*
   Speech act: Assertive.
   Theme: Personal satisfaction.
7. *Get some summer*.
   Speech act: Directive.
   Theme: Personal satisfaction.
8. *Have fun and taste beer*.
   Speech act: Directive.
   Theme: Personal satisfaction.
9. *Watch out*. *Love is in the air.*
   Speech act: Directive/ Assertive.
   Theme: Sensory experience.
10. *Enjoy refreshing taste in relaxing time.*
    Speech act: Directive.
    Theme: Personal satisfaction.

Dominant themes in French Beer Ads	**Appeals based on Hofstede's-Pollay's Framework**
Relaxation	Individualism
Pleasure	Pleasure
Personal Satisfaction	Individualism
Quality (of product)	Quality

Table 1: English Beer Ads

b. Translated Lebanese advertisements

1. في انا ما بنقطع هيدا جونا هيدا نحنا
   I'm the only who does not shut down (Unlike electricity). This reflects the atmosphere we live in.
   Speech act: Assertive.
   Theme: Continuity.
2. الليلة ليلة عيد و البيرة شكلها جديد
   Tonight is Christmas Eve and the beer has a new shape.
   Speech act: Assertive.
   Theme: Renewal.
3. بس اكبر بدي صير متلك
   When I grow up, I want to become like you.
   Speech act: Assertive.
   Theme: Family role models.
4. تنيناتنا للراشدين فقط
   Both of us are for adults only (Eyeglasses and beer).
   Speech act: Assertive.
   Theme: Wisdom.
5. مش معقول هينيكن 0.0 كحول
   Incredible. Heineken with 0.0 *alcohol.*
   Speech act: Assertive.
   Theme: Uniqueness.
6. اوعى خيك
   Take care of your brother. (National beer day)
   Speech act: Directive.
   Theme: Brotherhood.
7. ما حدا بغبر عليي
   No one can talk bad of me. (Leave dust on me)
   Speech act: Assertive.
   Theme: Reputation.
8. خبطة قدمكم
   The sound of your feet (Lebanese Independence Day).
   Speech act: Assertive.
   Theme: Patriotism.

9. صار وقتا
   It's about time.
   Speech act: Assertive.
   Theme: Season change.
10. هيدا قلبي... و مفتاحو
    This is my heart and its key.
    Speech act: Assertive.
    Theme: Romantic love.

Dominant themes in Lebanese Arabic Beer Ads	**Appeals based on Hofstede's-Pollay's Framework**
Family	Family
Quality (Beer)	Quality
Patriotism	Patriotism
Love (couple goals)	Collectivism
Wisdom	Wisdom

Table 2: Lebanese Beer Ads

c. Translated French advertisements

1. *Biere de Paris. C'est la meilleure.*
   Paris beer is the best.
   Speech act: Assertive.
   Theme: Taking pride in French beer.
2. *Biere Perle. La perle des bieres d'Alsace.*
   Pearl beer. The beer pearl of Alsace.
   Speech act: Assertive.
   Theme: Alsace fame for its beer quality.
3. *Enfin une grosse production. 100% francaise. Kronenbourg l'originale.*
   Finally a huge production. 100% French. The original Kronenbourg.
   Speech act: Assertive.
   Theme: Taking pride in French production.
4. *Actuellement en tournee dans toute la France. Kronenbourg l'originale.*
   Currently on tour throughout France. The original Kronenbourg.
   Speech act: Directive.
   Theme: Focus on pleasure and tourism in France.
5. *Face au meilleur la biere "33" export est prete a relever le defi.*
   Faced with the best "33" export beer is ready to meet the challenge
   Speech act: Assertive.
   Theme: Taking pride in the French beer production.
6. *La Grande Blonde de Champigneulles. La grande blonde a du corps.*
   The Tall Blonde of Champigneulles. The tall blonde has body.
   Speech act: Assertive.
   Theme: French femininity.

7. *La Grande Blonde de Champigneulles. Une grande blonde dans une bouteille.*
   The Tall Blonde of Champigneulles. The tall blonde in a bottle.
   Speech act: Assertive.
   Theme: French femininity.
8. 1664. *Une marque. Une empreinte*.
   1664. *A brand. A fingerprint*.
   Speech act: Assertive.
   Theme: Quality guarantee of French beer.
9. 1664. *Le gout a la francaise*.
   1664. *The French taste.*
   Speech act: Assertive.
   Theme: Quality of French beer.
10. 1664. *Quatre chiffres. Une biere*.
    1664. *Four letters. A beer*.
    Speech act: Assertive.
    Theme: Quality of French beer.

Dominant themes in French Beer Ads	**Appeals based on Hofstede's-Pollay's Framework**
Quality	Quality
Uniqueness	Uniqueness
Beauty	Beauty
Patriotism	Patriotism

Table 3: French Beer Ads

6. Discussion

The present study aimed to analyze qualitatively the speech acts that are present in Arabic, English, and French beer advertisements. For that purpose, thirty Lebanese Arabic, English, and French beer advertisements were collected and analyzed based on Searle's taxonomy of speech acts consisting of five categories: directives, assertives, commissives, expressives and declaratives.

The speech act analysis of English advertisements implies the dominant use of directives in order to address consumers. Directives usually request the audience to do something or take action. Hence, English advertisements use sometimes directives to encourage beer consumers to buy their products.

The content analysis of English beer advertisements revealed dominant themes like relaxation, pleasure, personal satisfaction and quality which compared against Hofstede's and Pollay's framework reflect individualism, pleasure and quality. This implies the importance of individualism as reflected in some ads in western culture. Hence English beer advertisements prioritize the life of the individual and their personal pleasure and satisfaction in general.

The speech act analysis of Lebanese Arabic advertisements indicates the use of assertives as means to address the audience. By implementing assertives, statements that are factual

are confirmed. This is how Lebanese promoters address their consumers. The content analysis of Lebanese beer advertisements shows that the dominant themes are family, quality, patriotism, love and wisdom equivalent to family, quality, patriotism, collectivism and wisdom in Hofstede's and Pollay's cultural framework.

Last but not least, the speech act analysis of French advertisements shows the use of assertives as means of persuading the consumers of the quality and uniqueness of French beer. The content analysis results imply the dominance of themes like quality, uniqueness, beauty and patriotism matching with Hofstede's-Pollay's cultural values.

As a conclusion, the overall analysis implies that English culture as reflected through the analyzed ads is more individualistic and focuses on self-satisfaction and relaxation as could be witnessed in beer ads. Lebanese and French ads share patriotism as a common theme since the focus is on the origins and traditions of the country. In the Lebanese case, the primary focus is given to festivities and national occasions like Christmas, independence day… whereas the French focus is more directed towards French regions, French originality and uniqueness. Collectivism seems to be more dominant in Lebanese beer ads as could be detected in themes such as wisdom, family, reunions and national occasions.

7. Conclusion

This study attempted to investigate similarities and differences between English, Lebanese, and French beer advertisements. Applying Searle's (1969) speech act theory and qualitative content analysis, the findings conveyed through the analysis of Lebanese, English, and French ads revealed that English beer advertisements implement directives while the Lebanese and French ones deploy assertives as dominant speech acts. The disparity also lies on the cultural level since English beer ads value individualism and self-satisfaction as dominant themes. French ads focus on country of origin, beauty and uniqueness which are relatively connected with patriotism whereas Lebanese beer ads focus primarily upon patriotism, national celebrations and collectivism as means to promote products. This connection between themes and cultural values shows the direct relation between culture and speech act usage. The choice of speech acts in beer ads seems to be directly influenced by them. When assertives are implemented, the focus shifts towards confirming collective values in order to appeal to the consumer whereas directives impel the person to move to action directly by stressing the individuality of the experience.

Last but not least, the study is an attempt to explore the interplay of speech acts and culture in English, French and Lebanese beer ads knowing that more samples could be collected in the future to cover a wider range of ads. However, what is evident is that cultural values differ from one country to another and their representation through the use of speech acts in beer ads confirms that assumption.

References

BLUM-KULKA, S. et al. 1989. *Requests and Apologies: A Cross Cultural Study of Speech Act Realization Patterns.* Jerusalem: Hebrew university.

CEMALOVIC, A. 2009. *Advertisements as Cultural Mirrors*. New York: Aarhus university.

CHENG, H. & SCHWEITZER, J. C. 1996. "Cultural values reflected in Chinese and U.S. television commercials". *Journal of Advertising Research* 36(3), 21–46.

CHILUWA, I. 2007. "A speech act analysis of written adverts of soft drinks in Nigeria". *Babcock Journal of Mass communication* 1(3), 29–38.

DAHL, S. 2004. *Cross-cultural Advertising Research: What do we Know about the Influence of Culture on Advertising*. London: Middlesex University Business School.

DYER, G. 1982. *Advertising as Communication*. London: Methuen.

ERLINGSSON, C. & BRYSIEWICZ, P. 2013. "Orientation among multiple truths: An introduction to qualitative research". *African Journal of Emergency Medicine* 3(2), 92–99.

GASS, S. & NEU, J. 2006. *Speech Acts across Culture: Challenges to Communication in a Second Language*. Berlin: Walter de Gruyter.

GASS, S. & SELINKER, L. 2008. *Second Language Acquisition: An Introductory Course*. London: Routledge.

HALL, E. T. & HALL, M. R. 1987. *Hidden Differences: Doing Business with the Japanese.* New York: Edward T. Hall.

HALLIDAY, M. A. K. 1978. *Language as Social Semiotic*. London: Arnold press.

HOFSTEDE, G. 1980. *Culture's Consequences: International Differences in Work-related value*. London: Sage.

HOFSTEDE, G. 2011. "Dimensionalizing Cultures: The Hofstede model". *Context Online Readings in Psychology and Culture,* 1–26.

HOFSTEDE, G. 2015 (June 12). *Culture*. Retrieved from Geert Hofstede: <http://geerthofstede.eu/culture>.

HOFSTEDE, G. 2015 (June 7). *National Culture*. Retrieved from The Hofstede Centre: <http://geert-hofstede.com/national-culture.html>.

HYMES, D. H. 1962. *The Ethnography of Speaking*. The Hague: Mouton.

KOTLER, P. 2000. *Marketing Management*. Englewood Cliffs, NJ: Prentice Hall.

MOOIJ, M. de. 2005. *Global Marketing and Advertising: Understanding Cultural Paradoxes*. London: SAGE.

MULLER, F. 2016. *What is a Speech Act? A Brief Introduction to Searle's Theory on Speech acts.* Grin Verlag.

NAUFALINA, L. 2017. "Speech act analysis of Bon appetite food advertisements". *Sastra Inggris-Quill* 6(6), 661–672.

POLLAY, R. 1983. "Measuring the cultural values manifest in advertising". *Current issues and Research in Advertising* 6:1, 71–92.

RAHMAN, M. 2008. *The Illocutionary Acts in the Movie Monster in Law*. Unpublished Thesis. State Islamic University of Alauddin Makassar.

SCHIFFRIN, D. 2003. *Approach to Discourse.* New York: Blackwell.

SEARLE, J. 1969. *Speech Acts*. Cambridge: Cambridge University Press.

SIMON, S. & DEJICA-CARTIS, D. 2015. *Speech Acts in Written Advertisements: Identification, Classification and Analysis.* Cambridge: Elsevier.

WIDYATAMA, R. 2007. *Pengantar Periklanan*. Yogyakarta: Pustaka.

Appendix 1: Pollay-Hofstede framework

Advertising Appeals	**Cultural Dimensions**
Individualism	
Freedom	
Uniqueness	
Collectivism	
Family	
Patriotism	
Popularity	
Beauty	
Social Status	
Sex	
Respect for the elderly	
Health	
Economy	
Nurturance	
Wisdom	
Neatness	
Quality	
Safety	
Security	
Adventure	
Magic	
Youth	
Competition	
Convenience	
Effectiveness	
Work	
Courtesy	
Enjoyment	
Leisure	
Natural	
Technology	
Modernity	
Tradition	

Appendix 2: English Advertisements

Fig. 1

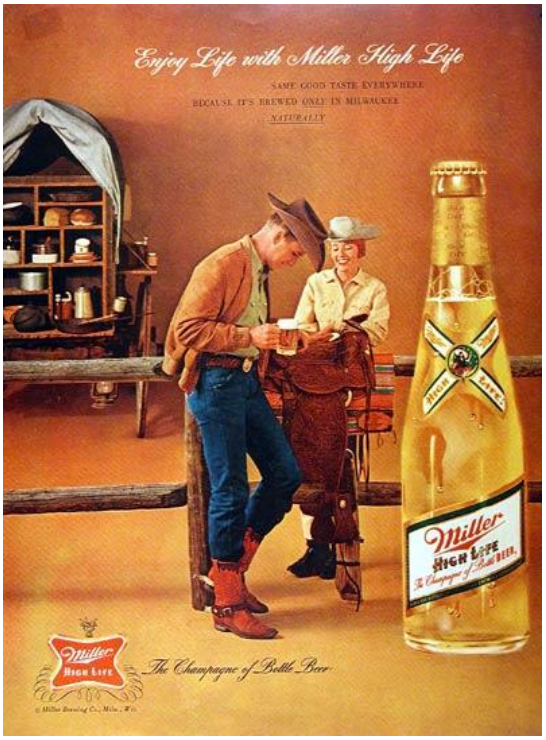

Fig. 2

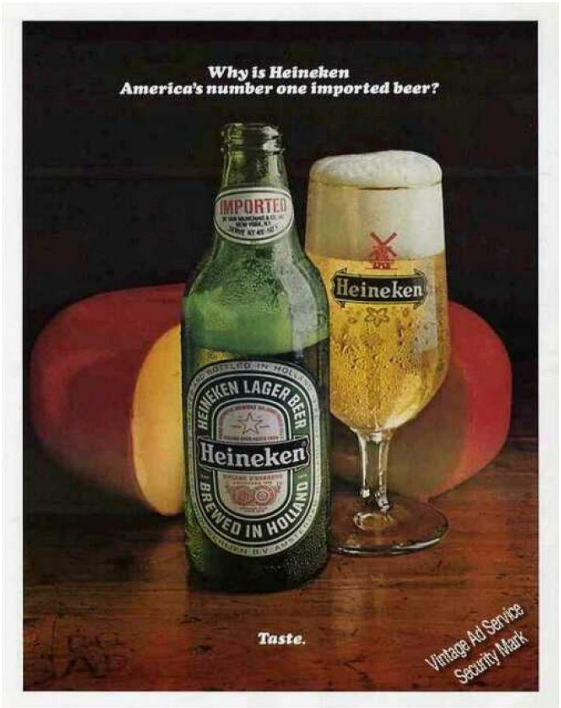

Fig. 3

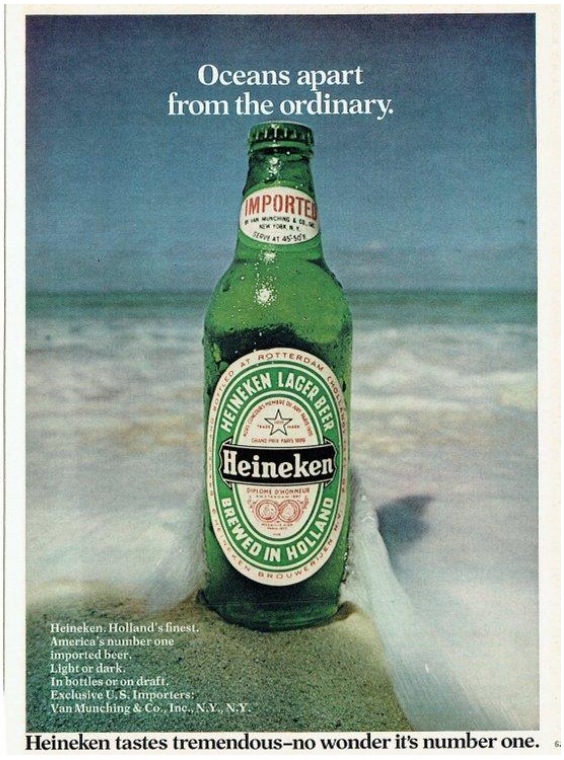

Fig. 4


IMPORTED
SERVE AT 45-50°F
HEINEKEN LAGER BEER
Heineken
BREWED IN HOLLAND
Heineken
Heineken tastes tremendous
IMPORTED HEINEKEN. IN BOTTLES, ON DRAFT AND DARK BEER.

Fig. 5


SHE FOUND
She Married Two Men
Budweiser
KING OF BEERS
"Where There's Life...There's Bud!"

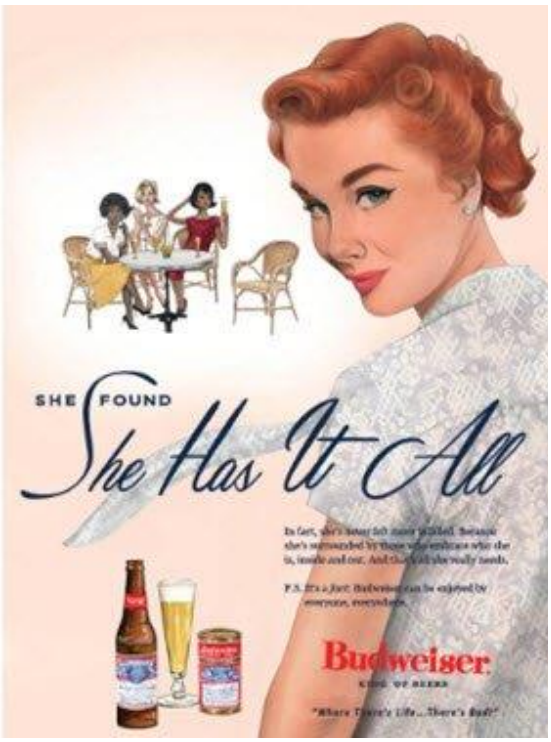
SHE FOUND
She Has It All
Budweiser
KING OF BEERS

Fig. 6

Fig. 7

Fig. 8

Fig. 9

Fig. 10

Appendix 3: Lebanese Advertisements

Fig. 1

Fig. 2

Fig. 3

Fig. 4

Fig. 5

Fig. 6

Fig. 7

Fig. 8


BEIRUT
BEIRUT
BEIRUT
صار وقتا
KASSATLY CHTAURA

Fig. 9


almaza
Pilsener Beer
هيدا قلبي...
ومفتاحو
هيدا جوّنا. هيدا نحنا.

Fig. 10

Appendix 4: French Advertisements

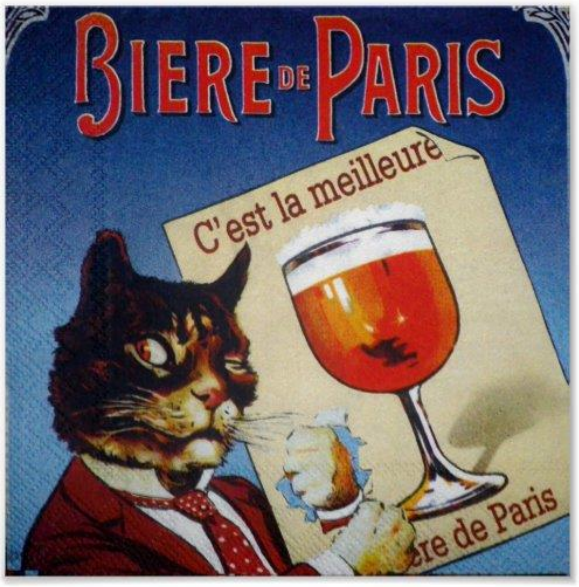

Fig. 1

Fig. 2

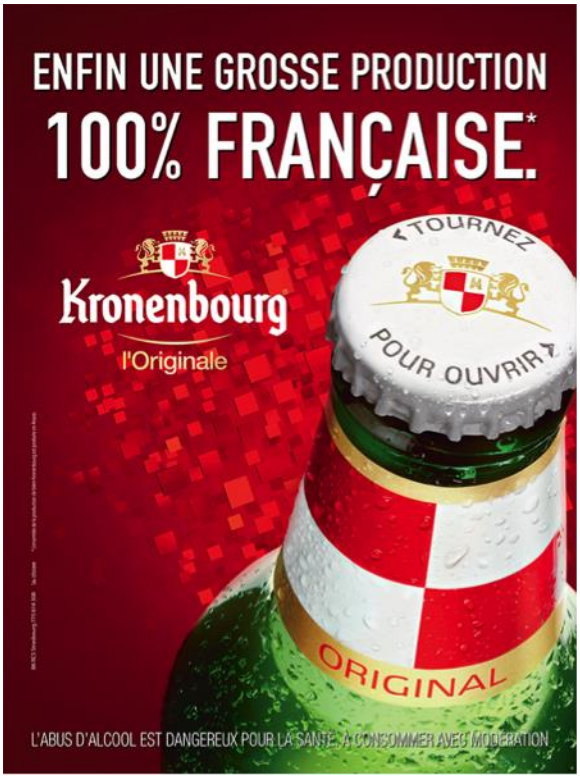

Fig. 3

Fig. 4


Face au meilleur
la bière "33" export
est prête à relever
le défi.

Fig. 5


la Grande Blonde de Champigneulles
LA GRANDE BLONDE
A DU CORPS...
Un corps plein et délicatement corsé,
une mousse fine et crémante,
une brillance ambrée.
Un corps moelleux,
une saveur finement amère
qui fait de la Grande Blonde de Champigneulles
une bière dont on reste gourmand.
la Grande Blonde de
Champigneulles

Fig. 6

Fig. 7

Fig. 8

Fig. 9

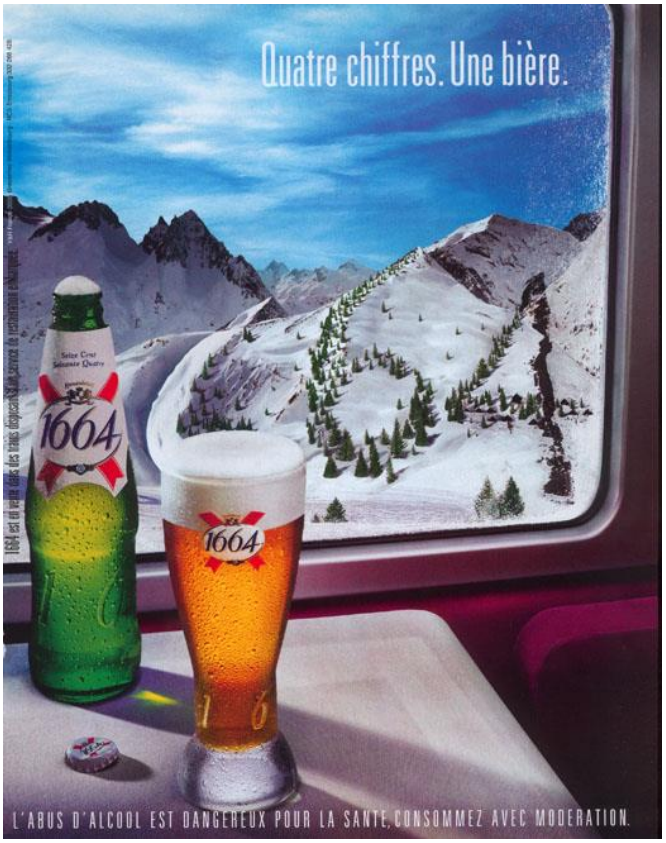

Fig. 10

Sources of advertisements

English Beer Ads

1. https://cdn.winsightmedia.com/platform/files/public/cspdn/600x450/mikes-hard-ad.png
2. https://encryptedtbn0.gstatic.com/images?q=tbn:ANd9GcSWM_TaPosX6jjCqFWg7jpantPkmXfQmMmAI-stsVv3YeihMdIN
3. https://s3.amazonaws.com/media.mediapost.com/dam/cropped/2019/03/08/bud-updated-ads-image_BTBUa7j_TbqlX1O.jpg
4. https://encryptedtbn0.gstatic.com/images?q=tbn:ANd9GcQiJEyaGqm47vcBDajyLOBuc7Grc6ODO1NT7qEczK0qGgzk0Ezb-w
5. https://encryptedtbn0.gstatic.com/images?q=tbn:ANd9GcQ1e7LThw0FUjnOMGAXL2b3XHwaqm8ikjSKKaMeMFEc7xajsbTj2w
6. https://i.pinimg.com/originals/a4/a3/7d/a4a37deed3e44c954d5bd981367bc3b9.jpg
7. http://www.adclassix.com/images/62millerbeer.jpg
8. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQVKHs-WJOBUO0JA0buKHMBLw-bww4DAj1yztt71nszNg7Z91Zl
9. https://i.pinimg.com/originals/01/89/9e/01899e014a2614818f50266ec84a40db.jpg
10. https://cdn.ebaumsworld.com/mediaFiles/picture/2183782/85021155.jpg

Lebanese Beer Ads

1. https://pbs.twimg.com/media/COZ_weMVAAA_KFp.jpg:large
2. https://www.lstatic.org/UserFiles/images/2017/divert/may2017/almaza-beirut-beer-ou3a-khayak.jpg
3. https://www.knooznet.com/upload/07-2019/article/large-1200x800%20(2).jpg
4. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcS-YBiPbP4_sZmtpx20dno_MFUsgeRykzjn4tyY1vX1V1aWh9kzNA
5. https://1.bp.blogspot.com/-bPq3iuBONvk/WhUSsvlVP2I/AAAAAAAAAqBI/Uyc-X6JAPuUQ7Y-_gPWTvh0UcO2YQPgwwCLcBGAs/s1600/Beirut%2BBeer.jpg
6. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRNZljy1p-W4eipAIh8bwL9UEHZLKGS4pdYZbJMXBYS9JotdDF3
7. https://pbs.twimg.com/media/CXAufi-WAAA0P9r.jpg
8. https://i.pinimg.com/originals/52/54/58/525458c2b2cd4d464aaa858d9aa2c5c7.jpg
9. http://arabadonline.com/mediafiles/articles/img-1479911683-4723.jpg
10. https://pbs.twimg.com/media/DHBVesRUMAAU5L_.jpg

https://encryptedtbn0.gstatic.com/images?q=tbn:ANd9GcSCBi8vkPXG1MEC_1xgJipAdyENdwZ7hj-xhOfpwwsyl-enAmY0

French Beer Ads

1. https://www.pinterest.com/happybento/vintage-beer-ads-posters/
2. https://www.zazzle.com/biere_de_paris_vintage_french_beer_ad_poster-228041609077369614
3. https://www.facebook.com/zeppelinhall/photos/new-draft-alert-kronenbourg-1664-from-france-kronenbourg-1664-is-frances-no1-sel/10153650134120880/
4. https://www.hot-mob.com/Showcases/kronenbourg-1664-carlsberg-street-party/
5. http://www.lebook.com/lacreative/creative/Pillier-ADVERTISING
6. https://picclick.fr/Publicit%C3%A9-Advertising-1969-Bi%C3%A8re-CHAMPIGNEULLES-la-181521590277.html
7. https://www.ebay.com/itm/M-Publicite-Advertising-1970-Biere-La-Grande-blonde-de-Champigneulles-/222855282382

Linguistica ONLINE. Issue 23. Published: May 7, 2020
http://www.phil.muni.cz/linguistica/art/dickins/dic-005.pdf
ISSN 1801-5336

AN ONTOLOGY FOR COLLOCATIONS, FORMULAIC SEQUENCES, MULTIWORD EXPRESSIONS, COMPOUNDS, PHRASAL VERBS, IDIOMS AND PROVERBS[*]1

James Dickins (University of Leeds, UK, J.Dickins@leeds.ac.uk)

Abstract: This article proposes an ontology (set of entities and explicit statement of the relations between them) for word-sequences (whether continuous or discontinuous) whose unifying feature is the co-occurrence within them of one or more of their words at greater frequency than would be predicted by their overall frequency of occurrence within the language. More precisely, the article proposes a number of possible ontologies, since, for some entities, it presents alternative possible definitions, discussing their merits and demerits. The article focuses almost entirely on English. It begins with a statement of general methodological principles. It argues we should not be attempting to discover what the true meaning of terms is, but to produce 'serviceable definitions' of terms which are at least relatively compatible with those produced by other writers and which can be coherently and explicitly related to other terms within the ontology. What I mean by a 'serviceable definition' is one which is of can be successfully used by researchers for the practical analysis of collocations and the other phenomena considered in this article. Whether the definitions proposed here are therefore 'serviceable' can only properly be judged by their successful deployment in future research.

The article then considers the following: collocations (Section 2), formulaic sequences (Section 3), multiword expressions (Section 4), compounds (sections 5-5.3), phrasal verbs (Section 6), idioms (Section 7) and proverbs (Section 8). Beyond these basic notions, the article considers other possible types of multiword expression (Section 9), further categories deriving from collocation, formulaic sequence and multiword expression (Section 10), semantic correlates of syntactic relationships in multiword expressions (Section 10), notions having fuzzy and discrete boundaries (Section 11), and universal and language-specific categories in the ontology (Section 12). Section 13 is a conclusion.

Key words: ontology, collocation, formulaic sequence, multiword expression, compound, phrasal verb, idiom, proverb

[*] Previously unpublished. [Editor's note]

[1] I thank Eric Atwell, Claire Brierley and two anonymous reviewers for *Linguistica* ONLINE for reading draft versions of this article and making very useful comments on it. These have considerably helped improve the final version. At various points in this final version of the article, I address comments made by the two *Linguistica* ONLINE reviewers on the earlier draft which they read, referring to them, where appropriate, as Reviewer 1 and Reviewer 2.

1. Introduction

This article has developed partly in response to the lack of clear global statements in the literature of the relationships between 'collocation' and a range of related notions.[2] Its aim, accordingly, is to drawn up an ontology (i.e. a set of entities and explicit statement of the relations between them) for such notions, in the hope that this will be useful for future researchers, who may either adopt it, or if they find it inadequate, build their own ontology covering the same or similar notions. The definition of 'ontology' adopted in this article is based on the fairly standard definition in logic, i.e. "set of entities presupposed by a theory" (Collins English Dictionary Online; henceforth CEDO). However, it recognises that a simple statement of the set of entities involved in a system does not tell us anything about how these entities relate to one another; hence the addition of the phrase, "explicit statement of the relations between them".

The article also takes the view that a lack of clear definitional statements is likely to lead to conceptual confusion, with different researchers unknowingly meaning different things by the same terms. The article does not base its view of collocations and related on a particular linguistic theory; and it is beyond its scope to consider collocations and related phenomena in relation to specific theoretical approaches (as done in Gries 2008, for example). The approach adopted is, however, intended to be commonsensical, and as such open to re-interpretation in terms of different theoretical approaches.

The article focuses almost entirely on English; in Section 13, it considers which of the notions which it puts forward may be considered universal, and which are specific to English (and perhaps some other languages). It also has the strictly limited aim of considering collocations, formulaic sequences, multiword expressions, compounds, phrasal verbs, idioms and proverbs from only the perspectives which specifically allow us to differentiate between them. These are mainly statistical, syntactic and semantic (the last largely treated from a purely denotative perspective). This means that many aspects of these notions which are essential in other respects are not discussed. Thus, in the case of idioms and proverbs, for example, I do not discuss issues such as (i) the relationship (e.g. metaphorical, metonymic, or other figurative relationship) between the idiom and proverb sense and the more basic ('literal') sense, (ii) the connotations which these may give to the idiom or proverb; (iii) the 'schematic' patterning of figurative idioms and proverbs (e.g. Kövecses 2010), (iv) the pragmatic and stylistic deployment of idioms and proverbs in different kinds of text, or (v) psycholinguistic issues (for a survey of all these aspects, see Gibbs 2010). These are all very important, but, as noted, inasmuch as they do not serve to differentiate the features which the article considers, they fall outside its scope.

Karl Popper has argued that in seeking to understand a term, we should not ask the question 'What is this really?', i.e. we should not attempt to search for the essence or 'true meaning' of that term – an approach which Popper (1986 [1957]: 26–43) calls methodological essentialism. Rather, we should attempt to provide what could be called a 'serviceable definition' of the term and use this definition as our starting point for the deployment of the concept (the defined term) in subsequent argumentation. To take a specific example, we

[2] This has proved a particular problem for a number of my doctoral research students doing corpus research in Arabic and Arabic-English translation.

should not ask 'What is capitalism really?'. Rather, we should start with a serviceable definition of capitalism, and then use this to investigate relevant phenomena in relation to this definition.

Popper's argument is based on common sense. It is possible to use a term such as 'capitalism' in many different ways, i.e. with many different definitions, whether explicit or implicit. (For a review of some of a number of different – and sometimes clearly incompatible – definitions of capitalism, see Merrill 1995.) The same situation is apparent in linguistics. There are thus, for example, numerous definitions for the term 'morpheme', many of which are clearly incompatible, and use 'morpheme' to mean quite different things for what is meant by the term in other approaches (for a survey, see Bauer 2004: 70–72).

Assuming various definitions of a term to be internally coherent and to apply sensibly to the facts to which they are relevant, there is no point in arguing which is the 'correct' definition. Indeed, this is counter-productive, since it deflects from seeing definitions as used within particular theories systematically, i.e. providing terms for concepts which fit into a systematic theoretical whole. Perhaps worse still, it encourages the epistemologically naïve view that there is necessarily a reality out there which it is our task as researchers simply to discover, rather than recognising the central importance of the theoretical approach which we adopt in shaping that reality.[3] This perspective echoes Saussure's dictum, "C'est le point de vue qui crée l'objet" (Saussure 1975 [1916]: 23), "It is the viewpoint which creates the object" (Saussure 1959: 8), or "it is the viewpoint adopted which creates the object" (Saussure 1983: 8).

There are, however, practical limitations to Popper's objection to methodological essentialism. Popper's views apply very well to abstract concepts, but are less applicable to concrete phenomena. Take the example of a rainbow. In an obvious sense, we all know what a rainbow is: we can point to one in the sky, we can describe what one looks like even if there is no rainbow to look at, we can draw one on a piece of paper. What we cannot, however, know without scientific investigation is the physics and particularly the optics which cause rainbows. In the case of a physical object like a rainbow, a form of methodological essentialism seems to be quite practicable. The basic definition of the phenomena – what they are, how they present themselves – seems obvious. What is of interest, rather, is what 'underlies' these phenomena analytically.

Physical phenomena such as rainbows and abstract concepts such as capitalism and the morpheme are extreme points on a continuum. We can fairly easily think of concepts which are more to the middle of this continuum. Examples include any semi-technical notion for which there is, however, fairly standard agreement among native speakers of a language about what is and is not included under the category concerned. An example might be the notion of a 'hobby'. Native speakers of English have a fairly good idea of what is and is not a hobby. It would be fairly perverse for a researcher to insist on a definition of 'hobby' which was different from that generally accepted by native speakers of English. If, however, there

[3] It might also be argued that it is also epistemologically naïve actively to deny the possibility that there is in fact a reality 'out there' which it is the task of researchers simply to discover. Between the 'absolute realist' and the 'instrumentalist' views, there is a more sophisticated overall perspective which accepts that we can never ultimately know whether what we describe as researchers is reality-as-it-really-is or a version of reality which presents itself as a result of the theory which we use to investigate that reality (cf. Mulder and Rastall 2005).

was no absolutely clear definition of what is and what is not a hobby among native speakers of English, a researcher could provide such a definition, basing themselves on the general views of native speakers, but adding specific criteria of differentiation (e.g. between a hobby and a sport) where this seemed to be necessary for clarity of definition.

In this article, I will consider terms of all three kinds discussed above: 1. Technical terms (of the 'capitalism' or 'morpheme' type), where it makes sense to provide a definition; 2. Semi-technical terms (of the 'hobby' type), where it makes sense to follow general usage, only introducing a new definitional element at the margins, to make plain precisely what is meant by the term in question; 3. Non-technical terms (of the 'rainbow' type), where there is clear existing general agreement about what the term refers to, and where there it would not be appropriate to try and provide a separate definition from this (what Lyons 1991: 32 has termed 'everyday metalanguage'). This division into three groups of terms is, of course, itself somewhat arbitrary. As noted above, there is, in fact, a continuum, such that terms may be more or less technical, or more or less 'everyday'. However, the division into three broad groups seems useful for practical purposes, provided that we remember that this division is a matter of convenience, and that boundaries between these types are, in reality, fuzzy (for further discussion of fuzzy boundaries, see Section 11).

The following are the central terms of these three different groups which I will consider in this article:

1. Technical terms, used in linguistics: collocations, formulaic sequences, multiword expressions;
2. Semi-technical terms, used in grammar teaching, etc: compounds, phrasal verbs;
3. Non-technical terms, in everyday usage (everyday metalanguage): idioms and proverbs.

As noted, these different groups of term require rather different treatment. Group 1, technical terms, are already defined in the literature in different and very often incompatible ways (and the notions they refer to also have alternative terms in some works). This makes it perfectly reasonable – and even necessary – to define them in specific ways for the purposes of this article, accepting that these definitions will necessarily not be compatible with all other definitions given in the literature. Such 'redefinition' (definition for the purposes of this article, and the underlying arguments it supports) should in practice, however, not involve the imposition of a meaning (definition) on the term involved which is so different from other previous definitions that it is likely to confuse readers who have already encountered these previous definitions. A fortiori, such redefinition should not be so far from previous definitions, that it appears a perverse usage of the term in question.

Group 2, semi-technical terms, already have fairly clear and fairly compatible definitions in the literature. Any 'redefinition' in these cases should, if it is to be sensible, only involve clarifying which of minor existing definitional differences this article adopts.

Group 3, non-technical terms, are better not redefined. It is possible to use an existing non-technical everyday term in a new technical sense intended to provide a more precise definition of what the non-technical term refers to. However, there are two points with this.

The first is a general problem that readers are in practice likely to confuse the new technical definition with the existing non-technical one, possibly even where the fact that the

term is being used in a specific technical sense is made plain in the article. The second point – and one which is clearly germane to this article – is that in using an existing non-technical everyday term, one may be intending to define, using technical notions, what is meant by this term in everyday language. This is the case in this article, which in its definitions of 'idiom' and 'proverb' is attempting to answer the questions: *Taking phenomena which are generally identified as idioms, how can we characterise/define idioms in technical linguistic terms?*; and *Taking phenomena which are generally identified as proverbs, how can we characterise/define proverbs in technical linguistic terms?* This is very different from Group 1, technical terms, where we are defining the terms for the purposes of this article, sometimes in ways which are clearly incompatible with the definitions of other writers. It is also different from Group 2, semi-technical terms, where we have at least some freedom to redefine terms.

2. Collocations

There are many different definitions of 'collocation' (cf. Firth 1957: 195; Cowie 1978: 132; Hausmann 1984; Richards, Platt and Webber 1985: 46; Sinclair 1991: 170; Kilgariff 1992: 29; Bahns and Eldaw 1993; Palmer 1993; Herbst 1996: 380; Hill 2000: 51; Lewis 2000: 132; Bartsch 2004: 68; Nesselhauf 2005; Seretan, 2011: 13. For summaries of these, see Bartsch Cai 2017: 4–7; Ruiz Yepes 2017: 11–18). In this article, the term 'collocation' will be used in a sense which is fairly standard in corpus linguistics to mean "a sequence of words or terms that co-occur more often than would be expected by chance within the context of a specific word" (Gómez 2009: 149).

Collocation in this sense is further explained by Lecheka (2015: 2):

> The strength of this kind of attraction between words can be measured through the statistical analysis of corpus data. The purpose of these statistical calculations is to find word pairs with significantly more co-occurrences than what would be expected by chance, given the words' total frequencies in the data. Thus, we can establish the most significant *collocates* of any given word in the language variety that the data represents […]

I will consider the inclusion of the word 'significantly' in Lecheka's characterisation of collocation below in this section. Before that, however, I will make two points in relation to Gómez's definition. Firstly, this definition does not impose a 'span' (or 'window'), i.e. a maximum number of words between a 'node' word (i.e. word of focal interest), and the other word(s) involved in the collocation. Secondly, it does not impose on a collocation that it should have any syntactic 'coherence', i.e. that the words involved in it should form some kind of syntactic unit, or sub-element within a syntactic unit. Thus, according to Gómez, words which occur at any distance apart from one another are technically collocations, provided they co-occur with greater frequency than would be expected by chance, given these words' total frequencies in the data. Similarly, while a form with syntactic coherence, such the noun phrase 'strong tea', involves a collocation of 'strong' and 'tea' under this definition, so does 'weak' and 'tea' in a form such as 'Do me some tea, but don't make it too weak'.

To operationalise 'collocation', i.e. to apply it in practice in a particular piece of research, it is necessary to do two things. The first is to specify a span (specific number of words before or after the node word). If one wanted a term to describe a collocation of this type, it could be called a 'span-defined collocation'. The second necessary step is to identify only those collocations which co-occur with *significantly* greater frequency than would be expected by chance, given these words' total frequencies in the data. This eliminates word co-occurrences which are technically collocations, but whose tendency to co-occur seems only insignificantly greater than would be predicted by their overall frequency of occurrence. In order to achieve this it is necessary to decide statistically what constitutes a significant (as opposed to insignificant) greater-than-chance frequency of word co-occurrence – a figure which would no doubt vary from study to study, depending on the specific goals of the study in question. A collocation of this type could be termed a 'statistically significant collocation' (or just a 'significant collocation'). A collocation which is both statistically significant and defined in terms of span could be called a 'span-defined (statistically) significant collocation'. There are different statistical approaches to collocations, which can yield quite different statistical significance results. A consideration of these falls outside the scope of this article (for discussion, see Evert 2007; Gries 2013).

3. Formulaic sequences

Various terms are used in roughly the same sense as 'formulaic sequence', e.g. 'formulaic expression', 'formulaic language' and 'prefab'; and various definitions have been given for these terms (cf. O'Donnell, Römer and Ellis 2013; Possio 2015: 60). Probably the best known definition is that of Wray and Perkins (2000: 1; also Wray 2002: 9), who define a formulaic sequence as "a sequence, continuous or discontinuous, of words or other elements which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use".

In this article, I shall adopt the following definition of formulaic sequence:

> A formulaic sequence is a collocation, whether continuous or discontinuous, which has syntactic coherence.

I propose this definition in preference to that of Wray and Perkins, in order to remove their notion of 'prefabricated', which has a technical psycholinguistic orientation, and could only be determined – if at all – by detailed psycholinguistic investigation. This is quite different from the current corpus-oriented approach, since corpora cannot directly tell us anything about what is stored and retrieved whole from memory.

Formulaic sequences which have a very high statistical occurrence are likely to occur in standardised contexts. Examples of such formulaic sequences are 'ladies and gentleman' (typically used at the beginning of speeches; cf. Mollin 2014: 149–151), and 'And they all lived happily ever after' (the traditional formula for ending a fairy story in English).

It might be argued that the current definition of 'formulaic sequence' is too broad for what is typically meant by 'formulaic', and that 'formulaic' implies, perhaps amongst other things, a high frequency of usage. If this were felt to be the case, it would be possible to add a further

criterion of statistical frequency to the definition of 'formulaic sequence' along the following lines:

> A formulaic sequence is a collocation, whether continuous or discontinuous, which has syntactic coherence and occurs with statistically greater frequency than a collocation which is not a formulaic sequence.

This would require defining, no doubt on a study-by-study basis, the precise statistical frequency which a syntactically coherent collocation would need to have in order for it to be (also) a formulaic sequence. The relationship between collocations and formulaic sequences can be represented as in Figure 1, which indicates that formulaic sequences are a subset of collocations; i.e. in linguistic-semantic terms, 'collocation' as defined in this article is a hyperonym (superordinate) of 'formulaic sequence'.

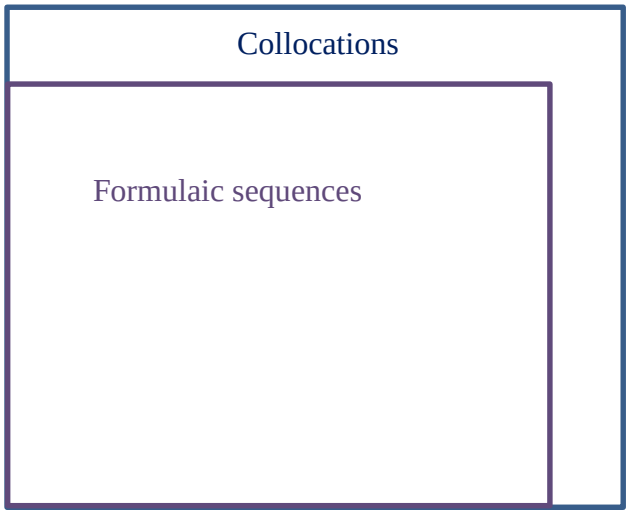

Figure 1: Semantic relationship between collocations and formulaic sequences

In addition to the criteria of syntactic coherence, and possibly 'significant frequency', used for defining 'formulaic sequence' above, we could also choose to regard 'formulaic sequence' rather differently – as a Group 2, semi-technical, term, or even a kind of Group 1, non-technical, term. In this case, we could classify collocations as non-formulaic or formulaic on the basis of native-speaker judgements, rather than on a syntactic (and also perhaps a statistical) basis. This proposed solution, of course, rests on native-speakers being able to make such judgements – and for the judgements which are made by different native speakers to be sufficiently similar that the results have an acceptably high degree of intersubjective acceptance across large numbers of native speakers. (This also raises the question of what constitutes an 'acceptably' high degree of intersubjective acceptance.)

A final redefinition of 'formulaic sequence' might combine aspects of a syntactic plus statistical definition with a native-speaker-judgement, definition. Thus, we could use statistical frequencies to determine 'potential' cases of formulaic sequences, and then use native-speaker judgements to decide whether identified potential cases are in fact to be regarded as cases of formulaic sequences. Having presented these three possible definitions of 'formulaic sequence', I will leave it open which one is the best to adopt.

4. Multiword expressions

The term 'multiword expression' is widely used, and has been defined in different ways. The following, taken from Constant et al. (2017: 840), are illustrative examples of different, and in some cases clearly incompatible, definitions: "a multiword unit or a collocation of words that co-occur together statistically more than chance" (Carpuat and Diab 2010: 242); "a sequence of words that acts as a single unit at some level of linguistic analysis" (Calzolari et al. 2002: 1934); "idiosyncratic interpretations that cross word boundaries" (Sag et al. 2002: 32); "lexical items that: (a) can be decomposed into multiple lexemes; and (b) display lexical, syntactic, semantic, pragmatic and/or statistical idiomaticity" (Baldwin and Kim 2010: 268). This last definition, with its use of the notion 'decomposed' draws explicitly on the notion of compositionality, i.e. the situation in which "the meaning of a complex expression is determined by the meanings of its constituent expressions and the rules used to combine them" (Wang 2018: 1), plus, we can add, the semantic correlates of these rules.

In this article I shall define a multiword expression, using the notion of compositionality, as follows:

> A multiword expression (MWE) is an expression consisting of two or more words, which is either: *Type 1*: fully non-compositional, i.e. none of the words has an independent sense; or *Type 2*: in which at least one of the words has a sense which is independent but is only found in the context of this expression; or *Type 3*: in which at least one of the words has a sense which is independent but is only found in definable limited contexts of which this context is one.[4]

It is possible to have a multiword expression which combines both Type 2 and Type 3, i.e. in which at least one of the words has a sense which is independent but this sense is only found in the context of this expression, *and* in which at least one of the words has a sense which is independent but only in definable limited contexts of which this context is one. Such an expression can be termed a *Type 2+3* multiword expression.

With regard to each constituent word of a multiword expression, this means that it may or may not have an independent sense, and that if it does have an independent sense this sense may be found in one context only, or in specific (i.e. definable) limited contexts, or in unlimited contexts. What is meant by 'context' here is more precisely termed 'lexical context', i.e. the context of another word-sense (another word in a specific sense). What is meant by 'unlimited context' is that there is no limit to the words (in particular senses; i.e. word-senses) in the context of which the word in question (in the particular sense in question) may be found. This does not mean that there is no limit to the meanings of other words in whose context the word (in the relevant sense) in question may standardly occur. Thus, the word 'court-martial' as a verb meaning 'try by court martial' (Oxford English Dictionary Online; henceforth OEDO); cf. 'court martial' as a noun meaning 'judicial court, consisting of military or naval officers, for the trial of military or naval offences, or the administration of martial law': OEDO) standardly has to have a word referring to a military institution as its

[4] For a theoretically explicit discussion of what is meant by 'independent sense' from the perspective of extended axiomatic functionalism, which I adopt elsewhere, see Dickins (1998: 241–244); see also the discussion of allosemic amalgamation in relation to morphology in Dickins (2006: 165, 188–189).

subject, whether this be a noun such as '[The, etc.] regiment', '[the, etc.] army', or '[the, etc.] military authorities', or a pronoun (e.g. 'he', 'they') which is co-referential with such a noun, or a proper noun referring to a person who holds an appropriate military rank (e.g. 'Kitchener'). These restrictions do not, however, constitute limited (lexical) contexts, since it is also perfectly possible to have other words as the subject of 'court-martial' which fall outside these categories. Thus, "Your cat court-martialled them" is a perfectly possible English sentence, however bizarre its meaning; and cf. the perfectly semantically reasonable "Your cat did not court-martial them, because cats cannot hold a relevant military rank". This contrasts with true limited (lexical) contexts where only specific words in specific senses are possible in the context of the given word in its specific sense, as discussed in this section below.

The situation with regard to the semantic independence of constituents in given (lexical) contexts can be diagrammatised as in Figure 2.

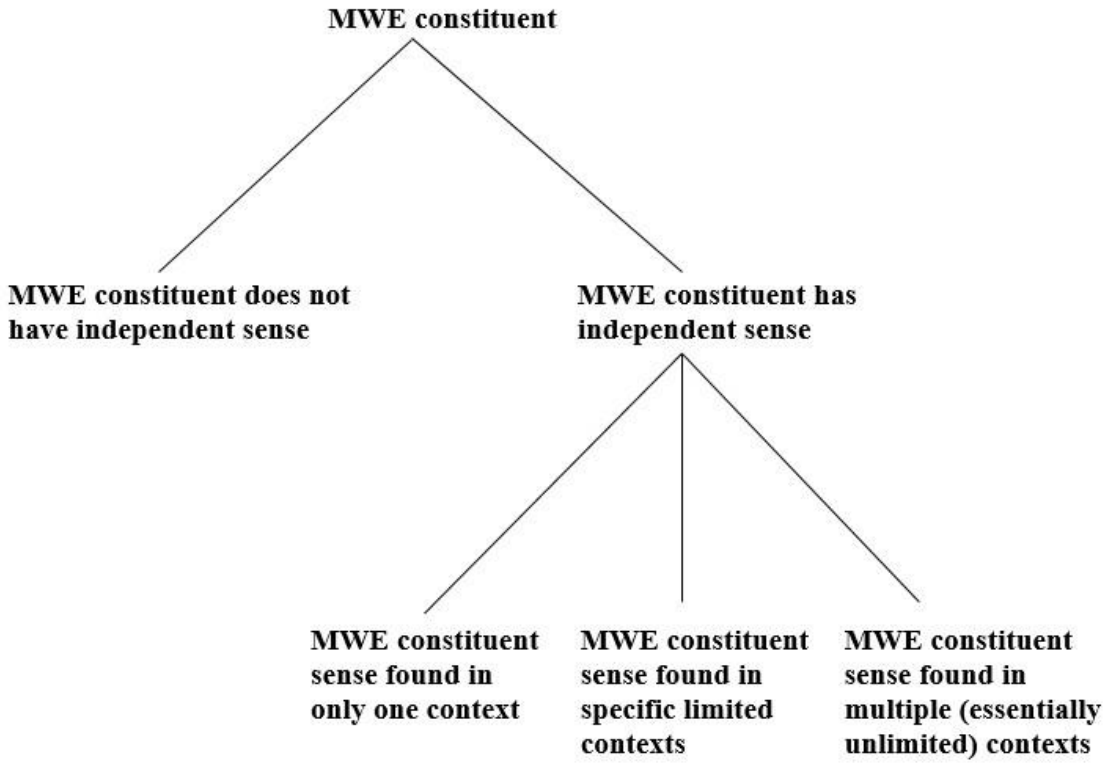

Figure 2: Typology of MWEs according to semantic independence of constituents

Semantic independence is closely related to the principle of compositionality (this section, above). In a compositional expression such as 'kick the football' the meaning of the whole expression is determined by the meaning of its individual expressions – i.e. the meaning of 'kick', the meaning of 'the' and the meaning of 'football' (in the relevant sense of all these words), plus the rules used to combine them: e.g. the fact that 'the football' is the object of 'kick', and that 'the' and 'football' together combine to make a noun phrase, and the semantic correlates of these 'rules' (syntactic relations). Each constituent in a phrase like 'kick the football' (i.e. 'kick', 'the' and 'football') can be said to have a 'free-compositional sense'; i.e. the sense which the constituent has in the complex expression in question, it also has in potentially unlimited other expressions. By contrast, a constituent having a sense which is

found in one context only or in specific limited contexts can be said to have a 'bound-compositional sense'.[5]

Probably most expressions in natural language are fully free-compositional (i.e. each of the words which makes them up has an independent sense). Multiword expressions, as defined in this article, are not. Consider the multiword expression (which is, of course, also an idiom; Section 7) 'kick the bucket', meaning 'die'. Here, it is impossible to say what sense each of 'kick', 'the' and 'bucket' has – because they do not in fact have separate (independent) senses. All we can say is that the entire phrase means 'die'. 'Kick the bucket' in the sense of 'die' is non-compositional. In fact, because none of the constituent words in 'kick the bucket' (= 'die') has a separate, independent sense, the expression is *fully non-compositional*, i.e. it is an example of a Type 1 multiword expression in terms of the definition of multiword expression given at the start of this section. The fact that none of the constituents of 'kick the bucket' has a separate, independent sense is also reflected in the fact that it is not possible to manipulate the multiword expression in any way: we cannot, for instance, say 'The bucket was kicked', or 'Don't go kicking any buckets, please'.[6]

In terms of the semantic independence of the constituents which make it up, 'kick', 'the' and 'bucket', 'kick the bucket' can be analysed as in Figure 3 (next page). Consider now, by contrast with 'kick the bucket', the multiword expressions (also multiword compounds; Section 5.1) *polar bear* meaning 'white carnivorous bear, Thalarctos maritimus […]': CEDO), and *brown bear* meaning 'large ferocious brownish bear, Ursus arctos […]': CEDO), i.e.

[5] A distinction needs to be made between a *constituent* which only occurs in one context, and a *constituent-sense* which only occurs in one context, i.e. a constituent having a sense which only occurs in one context. An example of a constituent which only occurs in one context is the morpheme 'cran', occurring only in the word 'cranberry' (a morpheme occurring in only one context being traditionally known as a 'unique morpheme', 'unique morph' or 'cranberry morpheme'; e.g. Carstairs-McCarthy 2002: 19). An example of a constituent-sense only occurring in one context is 'black' in 'blackbird' (cf. Section 5.2). A constituent only occurring in one context will either have a constituent-sense only occurring in one context or will not have an independent sense. An example of a constituent only occurring in one context having a constituent-sense only occurring in one context is 'cran' in 'cranberry'. Here the fact that 'berry' has a standard sense occurring in essentially unlimited other contexts forces us to conclude that 'cran' also has a sense here (cf. the analysis of 'blackbird' in Section 5.2), and since the constituent 'cran' only occurs this context, its sense also only occurs in this context. Examples of constituents only occurring in one context and not having an independent sense are 'spick' and 'span' in 'spick and span' (assuming 'span' not to be the same word as 'span', whose basic sense is 'interval, space or distance between two points': CEDO). ('And' in 'spick and span' also does not have an independent sense, though it does in unlimited other contexts.)

[6] There is one interesting apparent exception to the claim that 'kick the bucket' as a multiword expression cannot be manipulated in any way. This is the usage 'kick the proverbial bucket' (with 11 results on the IWeb corpus, henceforth IWeb, 24.9.18: <https://corpus.byu.edu/iweb/>). While 'proverbial' in this case formally goes with 'bucket', semantically, it relates to the whole phrase 'kick the bucket'. (Even more curiously, 'kick the bucket' is not a proverb, but a multiword expression, as seen, and also an idiom; Section 7.) Given the oddity of its semantics in relation to its syntax, 'proverbial' in 'kick the proverbial bucket' is not to be taken as the kind of manipulation of a multiword expression which demonstrates the semantic independence of one or more of its constituents. Other idioms similarly allow for the 'insertion' of the word 'proverbial', e.g. 'grasp the proverbial nettle'.

As Reviewer 1 has pointed out to me, it is in fact possible use forms of the type 'The bucket was kicked', or 'Don't go kicking any buckets, please' as semi-jocular 'transformations' of 'he kicked the bucket'. I have argued elsewhere that such forms fall outside the standard conventions of language, and as such are not to be considered in linguistic analysis (Dickins 1998: 324. The example I give there is 'I nearly strangled it', in 'Grasp the nettle!? – I nearly strangled it!').

both ‘polar bear’ and ‘brown bear’ are species of bear. It is possible to say things in English like ‘polar and brown bears’ (15 results on IWeb, 24.9.18: <https://corpus.byu.edu/iweb/>), or equally ‘brown and polar bears’ (20 results on IWeb: 24.9.18: <https://corpus.byu.edu/iweb/>). This demonstrates that ‘bear’ in both ‘polar bear’ and ‘brown bear’ has an independent sense and that this sense is the same in both compounds – the acceptability of forms such as ‘polar and brown bears’ and ‘brown and polar bears’ derives from the fact that ‘bear’ has the same sense in both ‘polar bear’ and ‘brown bear’ (on the basis of the general principles outlined in Cruse 1986: 49–83 for detailed argument of principles in this regard, some problematic cases notwithstanding). This is the sense which ‘bear’ has in multiple (unlimited) other contexts, ‘heavily-built, thick-furred plantigrade quadruped, of the genus *Ursus*’ (OEDO).

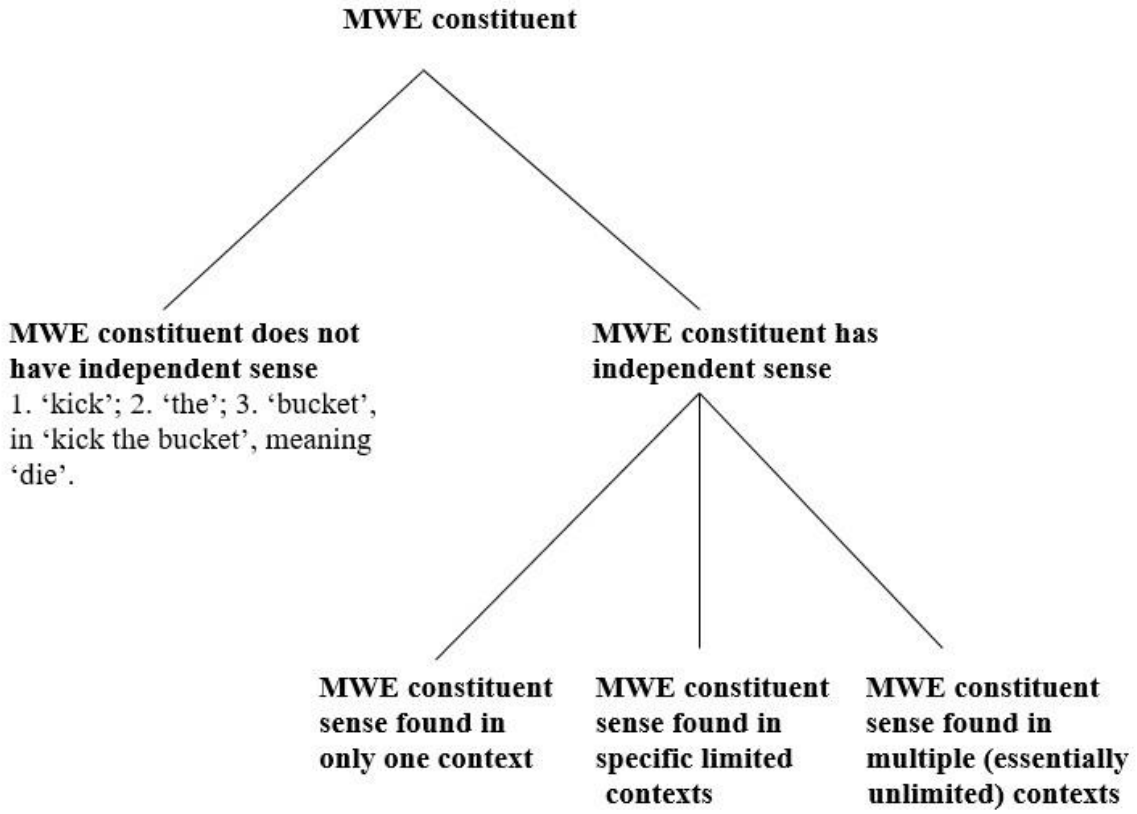

Figure 3: Analysis of the multiword expression ‘kick the bucket’ in terms of the semantic independence of their constituents

This, accordingly, is a Type 2 multiword expression, in terms of the definition of multiword expressions given at the start of this section, i.e. a multiword expression in which at least one of the constituent words has a sense which is independent but is only found in the context of this expression (in the case of ‘polar bear’, this word is ‘polar’, while in ‘brown bear’ it is ‘brown’).[7]

[7] I take it that in cases such as “bears/animals/those [etc.] of the ‘polar’ and ‘brown’ varieties”, the forms ‘polar’ and ‘brown’, with inverted commas around them, are correct; i.e. that this is a case of mention, rather than use, and that a form of this kind is not therefore a counterexample to the claim that ‘polar’ and ‘brown’ only occur in the contexts ‘polar bear’ and ‘brown bear’.

The conclusion that ‘bear’ in both ‘polar bear’ and ‘brown bear’ has the sense it has in multiple (unlimited) other contexts, ‘heavily-built, thick-furred plantigrade quadruped, of the genus *Ursus*’ requires us also to conclude that in ‘polar bear’, ‘polar’ (like the entire compound ‘polar bear’) has the sense ‘white carnivorous bear, Thalarctos maritimus […]’. It correspondingly requires us to conclude that in ‘brown bear’, ‘brown’ (like the entire compound ‘brown bear’) has the sense ‘large ferocious brownish bear, Ursus arctos […]’. ‘Bear’ is a hyperonym of ‘polar’ (also ‘polar bear’) in ‘polar bear’, and ‘bear’ is similarly a hyperonym of ‘brown’ (also ‘brown bear’) in ‘brown bear’.

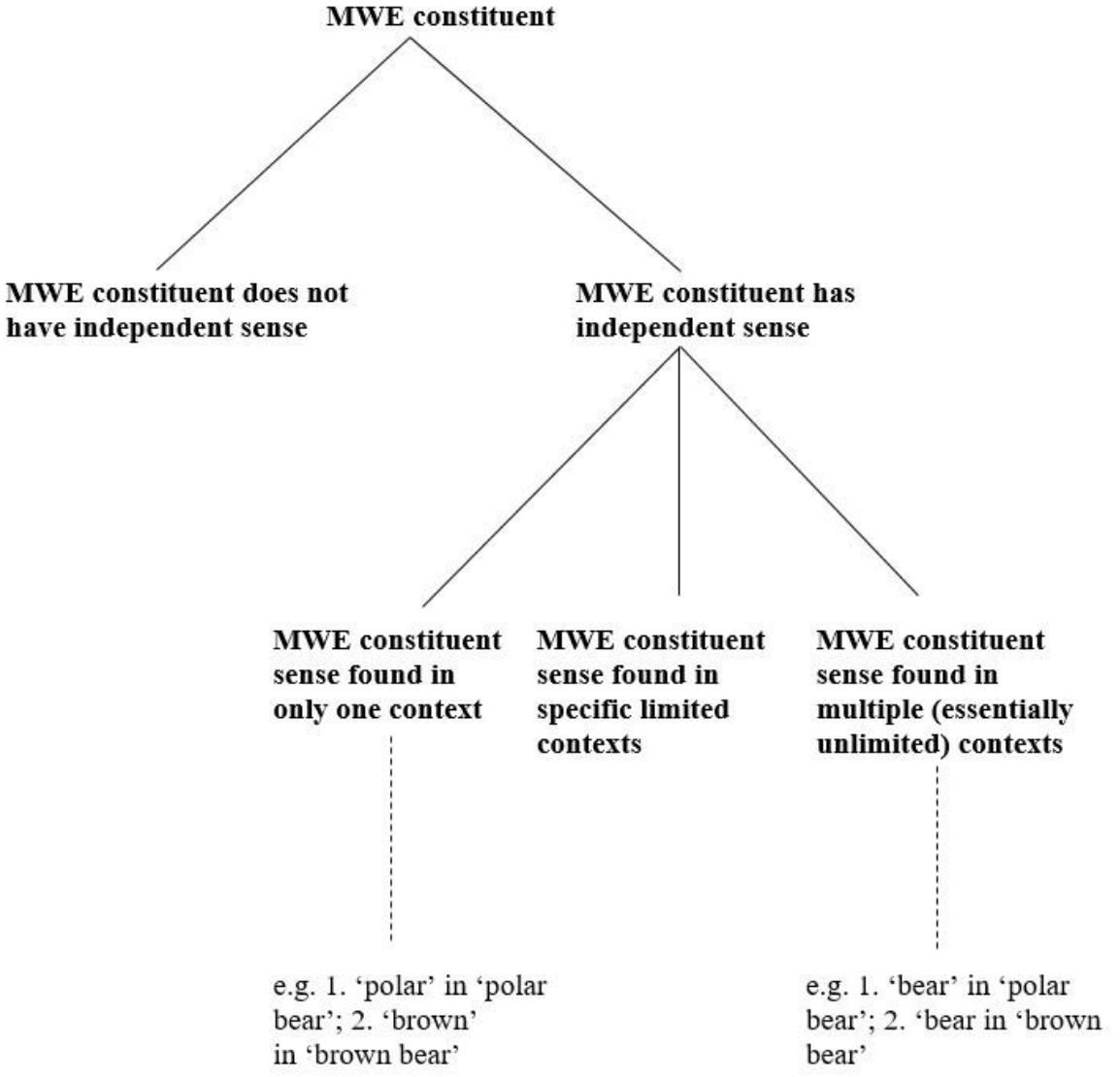

Figure 4: Analysis of the multiword expressions ‘polar bear’ and ‘brown bear’ in terms of the semantic independence of their constituents

This conclusion is supported by a consideration of the semantics of ‘grizzly bear’, i.e. ‘a variety of the brown bear, Ursus arctos horribilis’. The fact that ‘grizzly’ here means the same as it does in the entire compound ‘grizzly bear’ is shown by the fact that we can say ‘grizzly and polar bears’ (17 results on IWeb, 24.9.18: <https://corpus.byu.edu/iweb/>), i.e. by the same procedure which was used above to show that bear’ in both ‘polar bear’ and

'brown bear' has the same sense that it does in each of those compounds. In the case of 'grizzly bear', however (unlike that of 'polar bear' and 'brown bear'), this conclusion is secondarily demonstrated by the fact that it is possible to use 'grizzly' on its own, without a following 'bear', as a noun to mean the same as 'grizzly bear', e.g. 'I've just seen a grizzly'.

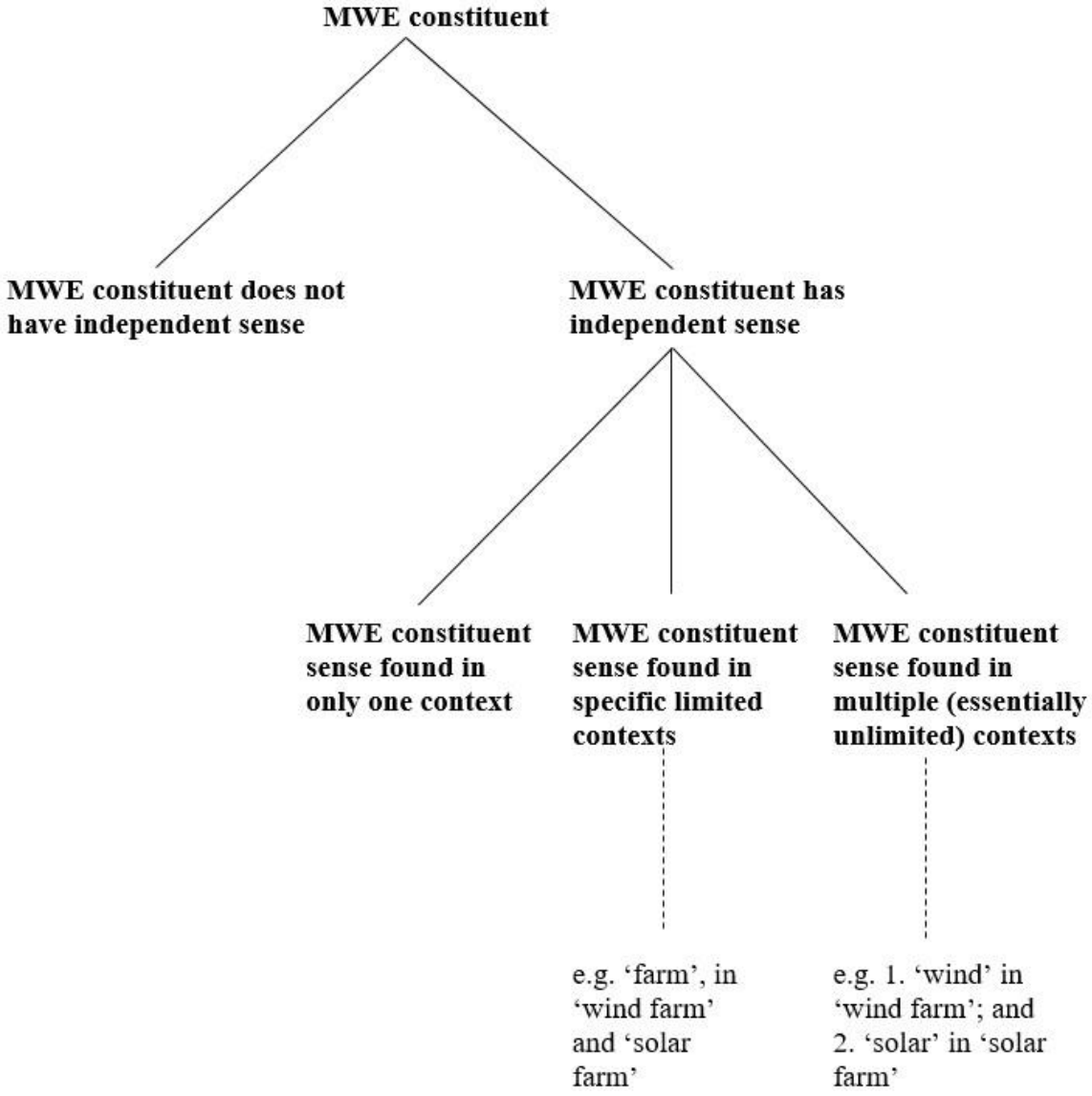

Figure 5: Analysis of the multiword expressions 'wind farm' and 'solar farm' in terms of the semantic independence of their constituents

In terms of the semantic independence of the constituents which make them up 'polar bear' and 'brown bear' can be analysed as in Figure 4 (previous page).

Examples of Type 3 multiword expressions, in which at least one of the words has a sense which is independent but only in specific limited contexts, are 'wind farm' and 'solar farm' (more precisely, these are multiword compounds: Section 5.1). Here, 'farm' has an independent sense, as can be seen from the fact that it is possible to things like 'wind and solar farms' (242 results on IWeb: 24.9.18: <https://corpus.byu.edu/iweb/>) and 'solar and wind farms' (101 results on IWeb, 24.9.18: <https://corpus.byu.edu/iweb/>). However, a form

such as ‘farms for wind energy’ does not appear possible (no results on IWeb, 24.9.18: <https://corpus.byu.edu/iweb/>). ‘Farm’ is thus only found in specific limited contexts in the (putative) sense ‘array of machinery for producing energy from a …. source’ (or similar), whereas ‘wind’ in ‘wind farm’ and ‘solar in ‘solar farm’ have a sense which is found in multiple (essentially unlimited) contexts.

In terms of the semantic independence of the constituents which make them up, ‘wind farm’ and ‘solar farm’ can be analysed as in Figure 5 (previous page).

As noted above, the definition which I have adopted in this article for multiword expression is different from that adopted by a number of other writers – and in fact rather narrower than that adopted by some. In particular, many writers, following Sag et al. (2002), extend the notion of ‘multiword expression’ to include expressions with only pragmatic and even statistical idiosyncrasies. Under the definitions adopted in this article, such expressions are analysed as collocations or formulaic sequences, rather than multiword expressions.

The relationship between collocations, formulaic sequences and multiword expressions can be represented as in Figure 6. This indicates that multiword expressions are a subset of formulaic sequences,[8] formulaic sequences themselves being, as noted (Section 3), a subset of collocations. In linguistic-semantic terms, ‘collocation’ as defined in this article is thus a hyperonym (superordinate) of ‘formulaic sequence’, which is a hyperonym (superordinate) of ‘multiword expression’.

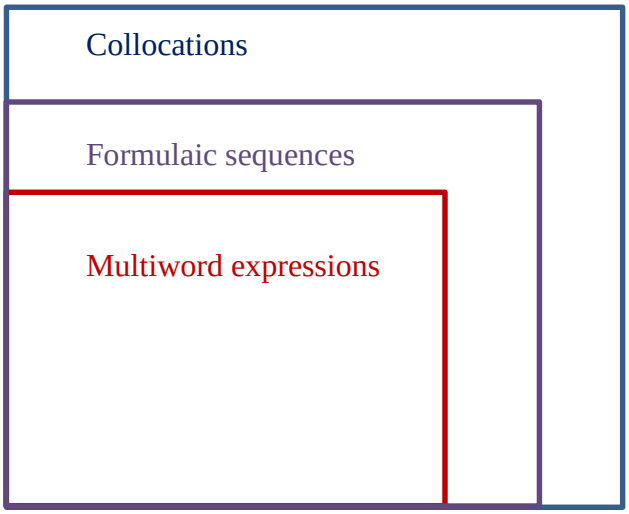

Figure 6: Semantic relationship between collocations, formulaic sequences and multiword expressions

5. Compounds

Compounds are expressions which may consist of one word, or two or more words and which contain at least one bound-compositional constituent (whether this occurs in only one context, or in specific limited contexts). A compound can be defined as follows:

[8] It would not be possible to regard multiword expressions as a subset of formulaic sequences if we were to include as part of the definition of the latter a relationship to specific situations or types of discourse. This is partly why I have chosen not to define ‘formulaic sequence’ in relation to situation/type of discourse in Section 3.

> A compound is a grammatical entity which consists of two or more elements each of which can appear as a separate word (in other contexts), and which is not fully free-compositional.

This definition is fairly compatible with standard definitions of 'compound'; e.g. "'the formation of a new lexeme by adjoining two or more lexemes" (Bauer 2003: 40), "A term used widely in descriptive linguistic studies to refer to a linguistic unit which is composed of elements that function independently in other circumstances" (Crystal 2008: 96).

Multiword compounds, i.e. compounds which consist of more than one word, are a type of multiword expression, while single-word compounds (compounds consisting of only one word) are not. Both multiword compounds and single-word compounds can be analysed in terms of the semantic independence of their constituents.

5.1 Multiword compounds

A multiword compound can be defined as follows:

> A multiword compound is a grammatical entity which is written (orthographically) as two or more words and consists of two or more elements, each of which can also appear as a separate word (in other contexts), and which is not fully free-compositional.

Consider in this respect, the examples 'sleeping policeman' meaning 'ramp in the road intended to jolt a moving motor vehicle, thereby encouraging motorists to reduce their speed' (OEDO),[9] 'wind farm', 'solar farm', 'polar bear' and 'brown bear' (these last four already discussed in Section 4, under multiword expressions). In terms of the semantic independence of their constituents, these examples can be analysed as in Figure 7 (next page).

As 'polar bear', 'brown bear', 'wind farm' and 'solar farm' have already been discussed in Section 4, in this section I will only consider 'sleeping policeman'.

[9] As Reviewer 2 has pointed out to me, there is an issue with how we determine that *sleeping* in *sleeping policeman* is a separate word, occurring also in other contexts, rather than something that just happens to be phonologically and orthographically the same as *sleeping* in *sleeping student* (i.e. one who sleeps). The same goes for *up* in *give up* in the case of phrasal verbs, and so on. It is often assumed in corpus linguistics that we just know this. As I have argued elsewhere, however, its determination, if it is to be rigorous, relies on a conception of what constitutes a word (or similar), and therefore ultimately on a particular theory of language (linguistic theory) within which the notion 'word' (or similar) is defined. This is an issue which I have addressed from the theoretical perspective of extended axiomatic functionalism in Dickins (1998: 227–240; also 187–198). One useful criterion (though not the only one) for claiming that two words – and by extension phrases – are the same is the presence of a figurative (e.g. metaphorical) relationship between them. This obtains, for instance, in the case of *sleeping policeman* in the sense of 'slumbering law enforcement officer' vs. *sleeping policeman* in the sense 'ramp in the road intended to jolt a moving motor vehicle, thereby encouraging motorists to reduce their speed' (cf. Dickins 2005; 2018), though it is not found with the same clarity, and in some cases not at all, in other examples discussed in this article. Given the theory-neutral orientation of this article, I assume that the analyses which I give of what do and do not consider to be words (etc.) are 'commonsensical', and leave it to the reader to decide whether they agree with these analyses or not.

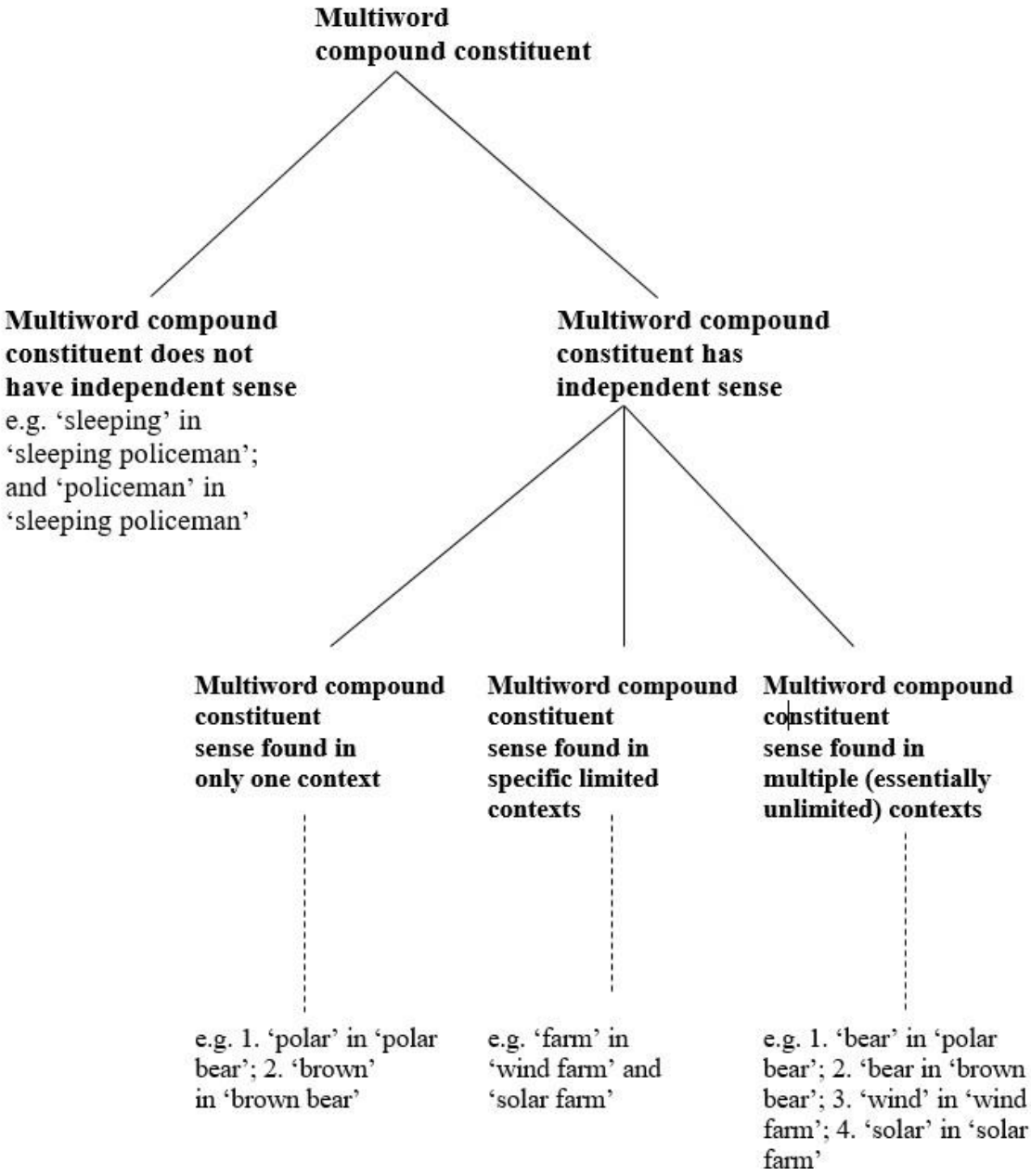

Figure 7: Analysis of multiword compounds according to semantic independence of constituents

Compounds involving constituents which do not have independent senses must logically have at least two such constituents – these two constituents (neither with an independent sense) together forming a larger constituent which does have independent sense. Thus the entire compound 'sleeping policeman' has an independent sense 'ramp in the road intended to jolt a moving motor vehicle [etc.]' (OEDO), but neither of the constituents which make it up, 'sleeping', or 'policeman', has an independent sense.[10]

[10] It is essential to distinguish between compositionality and semantic motivation. While neither of the elements 'sleeping' and 'policeman' in 'sleeping policeman' are semantically independent (the compound is semantically non-compositional), there is a clear metaphorical semantic motivation to the term 'sleeping policeman'.

5.2 Single-word compounds

A single-word compound can be defined, as follows:

> A single-word compound is a grammatical entity which is written (orthographically) as a single word and consists of two or more elements each of which can appear as a separate word (in other contexts), and which is not fully free-compositional.

Single-word compounds are by definition not a sub-type of multiword expression. In terms of the semantic independence of their constituents, however, single-word compounds can be analysed the same way as multiword expressions. This is illustrated in Figure 8 (next page) in relation to the following single-word compounds: 'ladybird' i.e. 'any of various small brightly coloured beetles of the family Coccinellidae' (CEDO), 'blackbird' 'common Eurasian thrush, Turdus merula, of which the male has black plumage and a yellow bill and is noted for its melodious song, and the female is dark brown' (OEDO), and 'trustful' and respectful'.[11]

'Ladybird' provides an example of a single-word compound in which the constituents do not have an independent sense. Like multiword compounds, single-word compounds involving constituents which do not have independent senses must logically have at least two such constituents – these two constituents (neither with an independent sense) together forming a larger constituent which does have independent sense. Thus the entire compound 'ladybird' has an independent sense 'Any of numerous small, domed beetles of the family Coccinellidae', but neither of the constituents (morphemes) which make it up, 'lady', or 'bird', does.

'Blackbird' provides an example of a second type of single-word compound constituent – one which has an independent sense, but this sense is only found in the context of this compound. Paralleling the analyses in Section 5.1 of 'polar' (in 'polar bear') and 'brown' (in 'brown bear'), in 'blackbird', both 'black' and the entire compound 'blackbird' have to be regarded as having the sense 'common Eurasian thrush, Turdus merula [etc.]'. I will discuss the ramifications of this analysis further below. For the moment, we should note, however, that 'black' in 'blackbird' does not have the sense "Of the darkest colour possible, that of soot, coal, the sky on a moonless night in open country, or a small hole in a hollow object; designating this colour; (also) so near this as to have no recognizable colour, very dark" (OEDO); i.e. 'black' in 'blackbird' does not have the standard colour sense of 'black'. This can be seen from the fact that not all blackbirds are black; in fact the female is brown (e.g.

Just as a policeman may do, a 'sleeping policeman' controls traffic speed, and like someone who is sleeping a 'sleeping policeman' does not actively intervene. Many of the examples analysed in this article (e.g. 'polar bear', 'brown bear', 'blackbird', 'wind farm', 'solar farm', are semantically motivated, in that they can be seen to be figuratively related (typically metaphorically related) to other more basic senses of the elements which they are made up of. Such semantic motivation, however, is entirely independent of the analysis of the semantic compositionality of the elements which make up these words and phrases.

[11] I have categorised 'trustful' and 'respectful' as (single-word) compounds, on the basis that they both consist of elements which can, in other contexts, occur as independent words – 'trust' and 'ful(l)' in the case of 'trustful' and 'respect' and 'ful(l)'. The fact that 'full' is written with two 'l's as a full word and one in this context is to be regarded simply as an idiosyncrasy of spelling. In terms of some definitions of 'compound', 'trustful' and 'respectful' are not compounds. However they are, in terms of the definition given at the start of this section.

https://www.rspb.org.uk/birds-and-wildlife/wildlife-guides/bird-a-z/blackbird/), and an albino blackbird would be white, while a blackbird which had fallen into a can of red paint would be red.

'Trustful' and 'respectful' provide an example of a third type of single-word compound constituent – one which has an independent sense which is found in more than one context (distinguishing it from the second type of single-word compound constituent above), but only in a limited number of contexts. In 'trustful' and 'respectful', the 'ful' constituent means 'having/exhibiting/showing' (or similar). This is not what 'full' means (in any of its senses) as an independent word; and 'ful' only has this sense in a specific number of words in which it occurs. (In many words in which 'ful' occurs, it does not have this sense, e.g. 'wonderful'.)

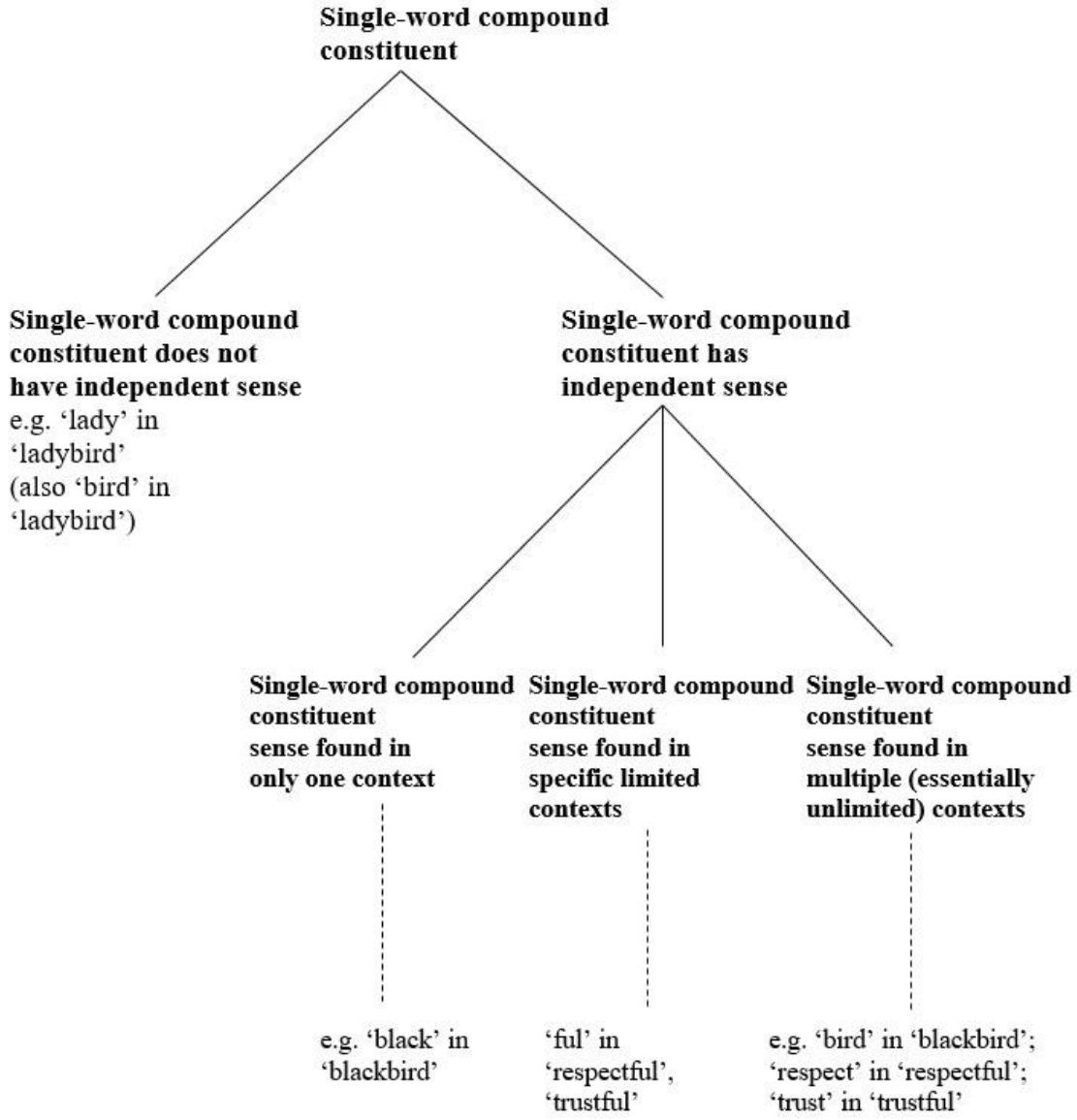

Figure 8: Analysis of single-word compounds consisting according to semantic independence of constituents

A final type of single-word compound constituent is one which has an independent sense, this sense being found in multiple (essentially unlimited) contexts. Examples are 'trust' in

'trustful' and 'respect' in 'respectful', where 'trust' and 'respect' mean what they mean (in the relevant sense) as independent words. Another, somewhat more subtle, example is 'bird', in 'blackbird'. Here, we should analyse 'bird' as having the sense 'Any feathered vertebrate animal: a member of the second class (*Aves*) of the great Vertebrate group […]' (OEDO), i.e. the same sense as it has in multiple (unlimited) other contexts. The reason for this is as follows.

Given, as argued above, that (i) the constituent (morpheme) 'black' in 'blackbird' has the same sense ('common Eurasian thrush, Turdus merula […]') as does the entire compound 'blackbird', and that (ii) 'bird' in 'blackbird' has the same sense which it has in multiple other contexts ('Any feathered vertebrate animal […]'), it follows that 'bird' in 'blackbird' is a hyperonym of 'black' in 'blackbird', as well, of course, of 'blackbird' itself. This proposed hyperonym-hyponym relationship can be represented as in Figure 9.

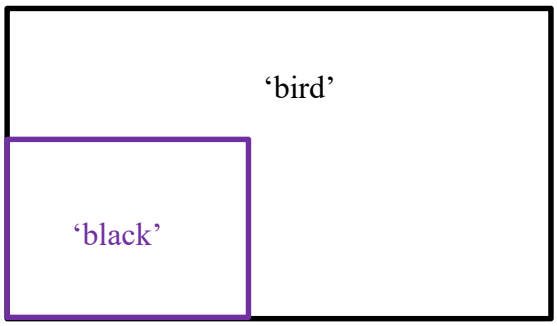

Figure 9: Hyperonym-hyponym relationship between 'bird' and 'black' (also 'blackbird') in 'blackbird'

Because 'black' (also 'blackbird') further delimits the sense of 'bird' here (as a hyponym of 'bird'), the fact that 'bird' (as the hyperonym/superordinate) has a wider sense than 'black' makes the element 'bird' irrelevant to the overall sense of 'blackbird'. This overall sense is simply defined by the sense of 'black' (a sense which, as noted, is only found in the context of 'blackbird').[12]

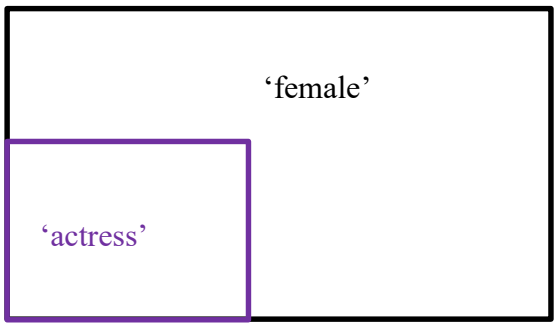

Figure 10: Hyperonym-hyponym relationship between 'female' and 'actress' in 'female actress'

[12] It should be noted that in Dickins (1998: 227–230; 433–435) I come to quite different conclusions regarding the semantic analysis of 'blackbird'. I now believe those conclusions to be wrong.

The reasonableness of this argument can be seen by considering a form such as ‘female actress’, e.g. as in ‘she’s a female actress’, where ‘female’ is a hyperonym and ‘actress’ a hyponym: all actresses are female, but not all females are actresses. The fact that ‘female (as the hyperonym) has a wider sense than ‘actress’ makes it irrelevant to the overall sense of ‘female actress’. This overall sense is simply defined by the sense of ‘actress’. This can be represented as in Figure 10 (previous page), paralleling Figure 9.

The analysis of ‘blackbird’ in this section parallels that of ‘polar bear’ and ‘brown bear’ in Section 5.1, where it was argued that ‘polar’ and ‘brown’ are hyponyms of ‘bear’, i.e. that ‘polar’ and ‘brown’ independently denote specific types of bears. The general proposal that compounds may consist of hyperonym-hyponym pairs in which the overall compound is also a hyponym of one of the constituents making up the compound was supported for multiword compounds by the analysis in Section 5.1 of ‘grizzly bear’.

Figure 11 shows the semantic relationships between collocations, formulaic sequences (formulaic language), multiword expressions, and compounds.

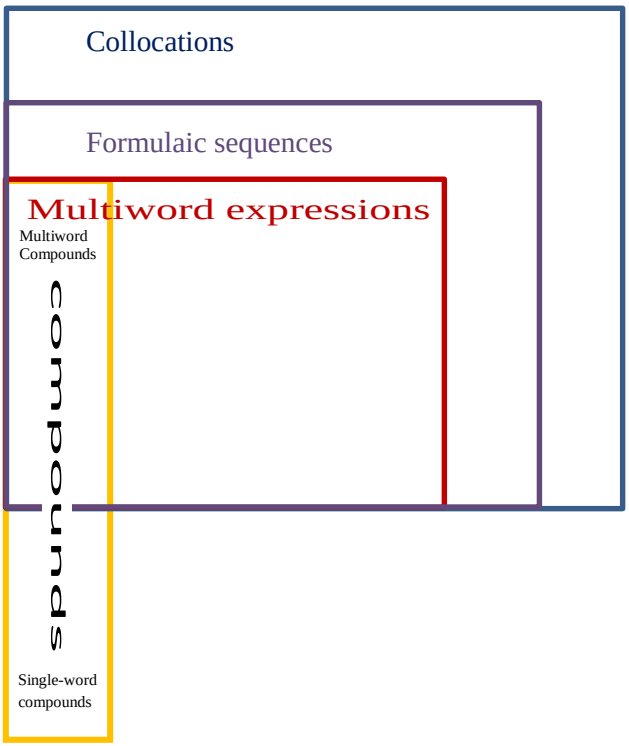

Figure 11: Semantic relationship between collocations, formulaic sequences, multiword expressions, and compounds

As noted above, while multiword compounds are a sub-type (subset) of multiword expressions, single-word compounds, by definition, are not. We should also acknowledge the fairly ad hoc nature of the distinction between multiword and single-word compounds. In many case, a compound can be written as a single word, or two words, or as two words with a hyphen between them; e.g. ‘desertsurfing’, ‘desert surfing’, and ‘desert-surfing’.

5.3 Some issues with the definition of 'multiword compound'

It might be felt that the definition of 'multiword compound' given above is too restrictive, as it excludes all forms which are fully free-compositional. There are two other alternative definitions of compound, therefore, which might be felt to be closer to standard usages of 'compound' in linguistics, which I will consider here.

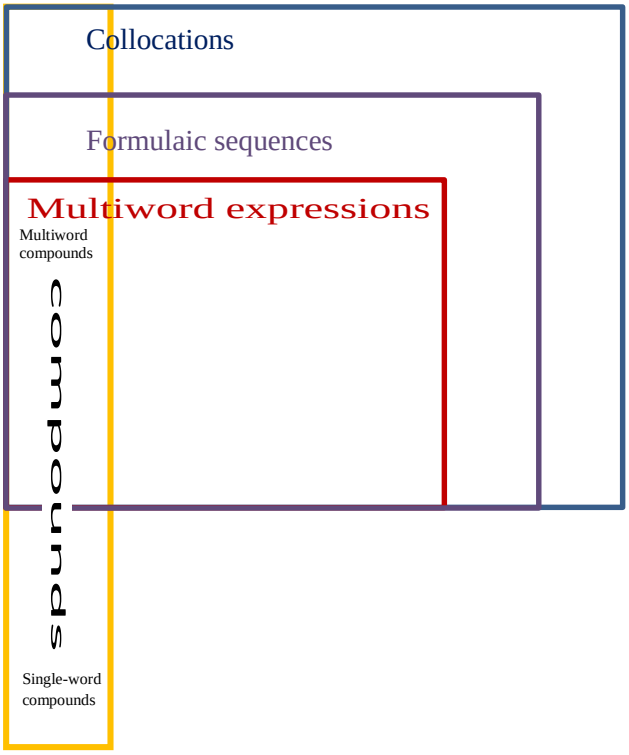

Figure 12: Semantic relationship between collocations, formulaic sequences, multiword expressions, and compounds defined to include all collocations

The first is to extend the notion of 'multiword compound' to include collocations of the grammatically appropriate kind (and therefore, by definition, also include formulaic sequences of the grammatically appropriate kind). This would give a scope for multiword compound (and compounds more generally) as in Figure 12. According to this definition, compounds would include any relevant form in which words co-occur with a greater frequency than is predicted from the overall frequency of occurrence of the words in isolation. Thus a commonly occurring collocation such as 'door key' (765 results on IWeb: 26.9.18: <https://corpus.byu.edu/iweb>) would count as a compound under this approach, notwithstanding the fact that it is fully free-compositional (cf. Section 11). By contrast a form such as 'cave key' (e.g. 'key for opening up a cave, via opening an iron door with railings which has been placed in front of it' (21 results on IWeb: 26.9.18: <https://corpus.byu.edu/iweb>), which we can take to not be a collocation, would be excluded. 'Compound' as defined in this way is represented in Figure 12. (Overall, 'door' occurs 1,561,423 on the c. 1.4 billion-word IWeb corpus, while 'cave' occurs 199,087 times; i.e. 'door' occurs around 7.84 times more frequently than 'cave'. However, 'door key' with 765 occurrences occurs around 36.43 times more frequently than 'cave key' with 21 occurrences. 15 of the 21 occurrences of 'cave

key' occur in the walkthrough for the computer game 'Turok 2: Seeds of Evil': <http://www.the-spoiler.com/ACTION/Acclaim/turok.2.1.html>).

A further extension to the definition of 'compound' would be to treat any form which is of the grammatically appropriate kind as a compound, regardless of whether it involves a collocation or not. Under this definition, 'cave key' as well as 'door key' would count as a compound. 'Compound' under this definition can be represented as in Figure 13.

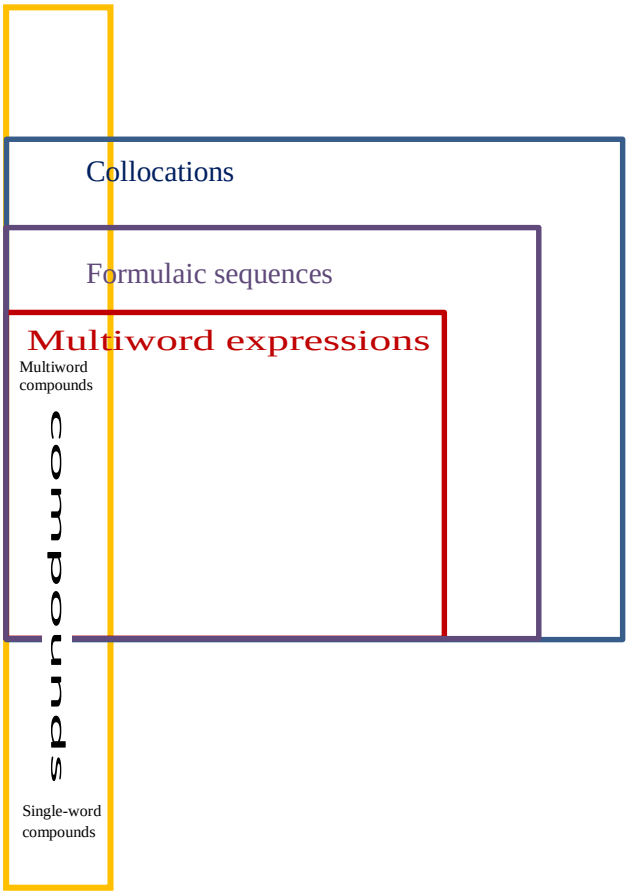

Figure 13: Semantic relationship between collocations, formulaic sequences, multiword expressions, and compounds defined to include all grammatically appropriate word co-occurrences

Under this definition, the notion 'compound' simply refers to specific set of grammatical structures. Since multiword compounds under this definition extend beyond the scope of collocations, the notion of 'compound' is not relevant to defining phenomena which fall within the scope of the notion of 'collocation', which is the basic concern of this article. Accordingly, 'compound' under the definition represented in Figure 13 is not a useful notion for this article.

There is, however, a further rider to this. If we decide in linguistics 'compound' is standardly used to describe a set of purely grammatical entities, we may be best advised to agree to use 'compound' in this way. In this case, we could use the term 'MWE-compound' to refer to what has been earlier termed a 'compound' in this article (e.g. in Figure 12). Other terms would also be available, such as 'formulaic compound' to describe any relevant form

which also falls within the category of formulaic sequence, and 'collocational compound' to describe any relevant form which also falls within the category of collocation.

Having raised these possibilities, I will in the remainder of this article continue to use 'compound' as represented in Figure 12, i.e. to refer only to appropriate grammatical structures which are also multiword expressions or not fully free-compositional single-word compounds.

6. Phrasal verbs

A phrasal verb can be defined as follows:

> A phrasal verb is a phrase that consists of a verb with a preposition or adverb or both, and is not fully free-compositional.

This is fairly close to definitions of phrasal verb used elsewhere, e.g. "a phrase that consists of a verb with a preposition or adverb or both, the meaning of which is different from the meaning of its separate parts" (Cambridge Dictionary Online). A phrasal verb, under the definition given here, is a type (sub-type) of multiword expression. Under this definition, if something is classified as a phrasal verb, it is not also to be classified as a multiword compound; i.e. 'multiword compound' and 'phrasal verb' are defined in this article as disjunct (non-overlapping) classes.

Examples of phrasal verbs are 'give up' (= 'leave off; to cease from effort, leave off trying; to stop': OEDO), 'fish out' (= 'find or extract'), 'turn on' (= 'to excite, interest, fill with enthusiasm; to intoxicate with drugs, to introduce to drugs; to arouse sexually': OEDO, i.e. the antonym of 'turn off').

Phrasal verbs can be analysed according to the semantic independence of their constituents as in Figure 14 (next page).

'Give up' is an example of a phrasal verb in which neither of the individual constituent words has an independent sense.

'Fish out', meaning 'find or extract' (as in "The supervisor's role is to fish out any old candle wicks using the small sticks if the wax is from old candle remnants": IWeb: <http://www.youthwork-practice.com/ideas-kids-crafts/Candle-Making.html>), provides an example of a second type of phrasal-verb constituent. Here 'fish' has an independent sense, but this sense is only found in the context of this phrasal verb. 'On' and 'off' in 'turn on' (= 'excite, interest, fill with enthusiasm', etc.), and 'turn off' (= 'put (a person) off, repel, disillusion, cause to lose interest', etc.) provide further examples of phrasal-verb constituents having independent senses which are only found in the context of a particular phrasal verb. The word 'turn' here can be glossed as meaning '(un)excite, (dis)interest', i.e. as subsuming semantically (being the hyperonym/superordinate of) both the active emotion of 'excite/interest' and the passive one of '*un*excite/disinterest'. The fact that 'turn' has an independent sense in 'turn on' (= 'excite, interest, fill with enthusiasm', etc.), and 'turn off' ('put (a person) off, repel, disillusion, cause to lose interest', etc.) can be seen from the acceptability of things like 'turns you on or off' (5 occurrences in the relevant sense on IWeb, 24.9.18: <https://corpus.byu.edu/iweb/>), e.g. "When you see other presenters, notice what they say

that turns you on or off. Adapt what you learn to your own presentations and style": <https://www.earlytorise.com/zen-and-the-art-of-speaking-at-seminars/>. The fact that 'turn' has an independent sense in the phrases 'turn on' and 'turn off' in the relevant senses further requires us to conclude that so do 'on' and 'off'. This, however, does not seem to be a sense which occurs elsewhere. Thus, we cannot use 'on' on its own to mean 'excited, interested', etc. or 'off' on its own to mean 'unexcited, disinterested', etc. Nor do there seem to be other multiword expressions in which 'on' and 'off' mean 'excited, interested', etc. and 'unexcited, disinterested', etc. 'On' and 'off' in 'turn on' and 'turn off', in the relevant senses, thus only occur in the independent sense which they have here in the context of 'turn'.

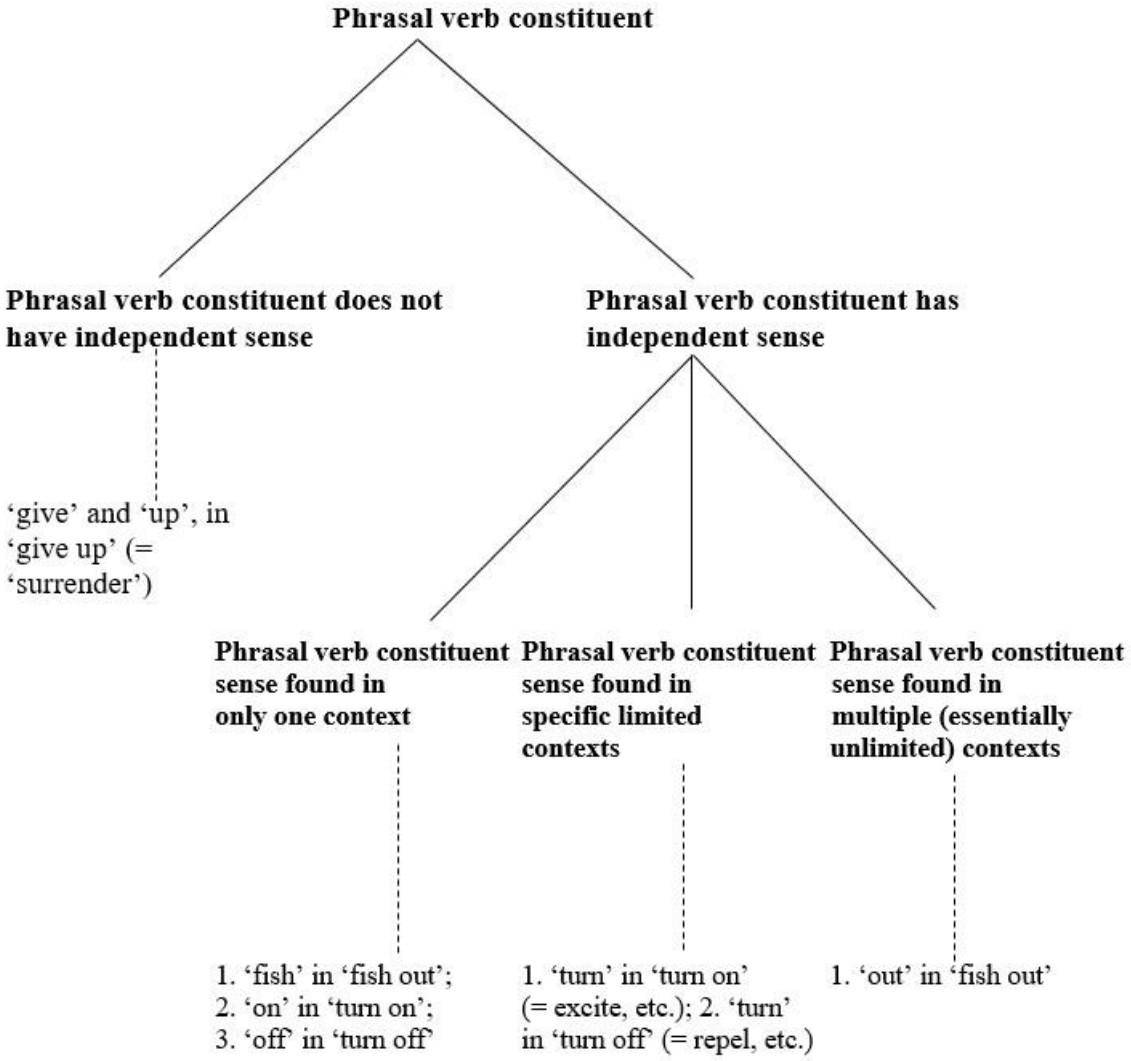

Figure 14: Analysis of phrasal verbs according to semantic independence of constituents

'Turn', meaning '(un)excite, (dis)interest', etc. in 'turn on/off' (in the relevant senses) is an example of a phrasal verb constituent having an independent sense occurring only in limited contexts. 'Turn' seems in fact to only occur with this sense in the two contexts of 'on' and 'off'. There do not seem to be any other multiword expressions in which it has the same sense.

A final type of phrasal-verb constituent is one which has an independent sense, this sense being found in multiple (essentially unlimited) contexts. An examples is 'out' in 'fish out' (= 'find, extract'). Thus, 'out' can occur in the sense which it has in 'fish out' (roughly 'in/to [the] outside of') in essentially unlimited contexts, for example with verbs such as 'walk out' (e.g. 'he walked out of the room'), 'pull' (e.g. 'she pulled the hamster out of the hole'), or even without a verb, e.g. 'out of there, please!'.[13]

Figure 15 shows the semantic relationships between collocations, formulaic sequences, multiword expressions, compounds and phrasal verbs.

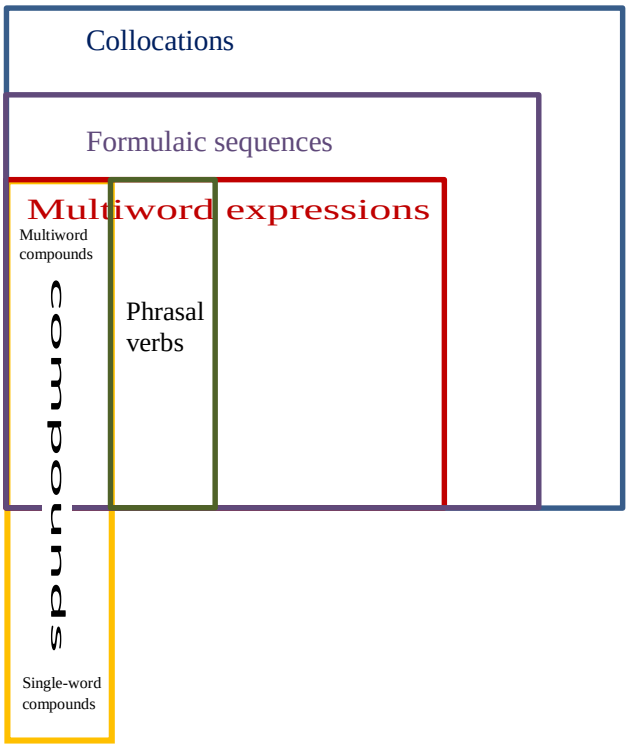

Figure 15: Semantic relationship between collocations, formulaic sequences, multiword expressions, and compounds and phrasal verbs

7. Idioms

An idiom is a particular kind of multiword expression. Precisely defining idioms has, however, proved problematic. For attempts, see Grant (2003), Grant and Bauer (2004), Liu (2008) and Wulff (2008). I will define 'idiom' as follows:

> An idiom is a phrase that is not a compound, is not a phrasal verb, is non-clausal, and is not fully free-compositional.

[13] The account of the semantics of phrasal verbs which I have given here is not complete (see also Section 1). In addition to issues of the semantic independence of the components of phrasal verbs, which involve denotative meaning, there are also issues of connotative meaning, most prominently what Hervey and Higgins term 'reflected meaning' (e.g. Dickins, Hervey and Higgins 2017: 103-104), and, associated with this, metaphor (cf. Dickins, Hervey and Higgins 2017: 194-210, and, rather more rigorously, Dickins 2005; Dickins 2018).

This definition excludes compounds (Section 5), phrasal verbs (Section 6) and also proverbs, which are clausal (Section 8). This is an attempt to mirror what is generally meant by idioms in everyday language (idioms being classified as non-technical in Section 1). Although most idioms in English are figurative (mainly metaphorical, but also sometimes metonymic, etc.), figurativeness is not a defining feature of idioms. Thus '(as) sure as eggs is eggs' meaning 'absolutely sure' is an idiom, but is not figurative, as is 'by and large' (assuming we classify this as an idiom, rather than a multiword compound; this section, below).

In terms of the semantic independence of their constituents, idioms can be analysed as illustrated in Figure 16.

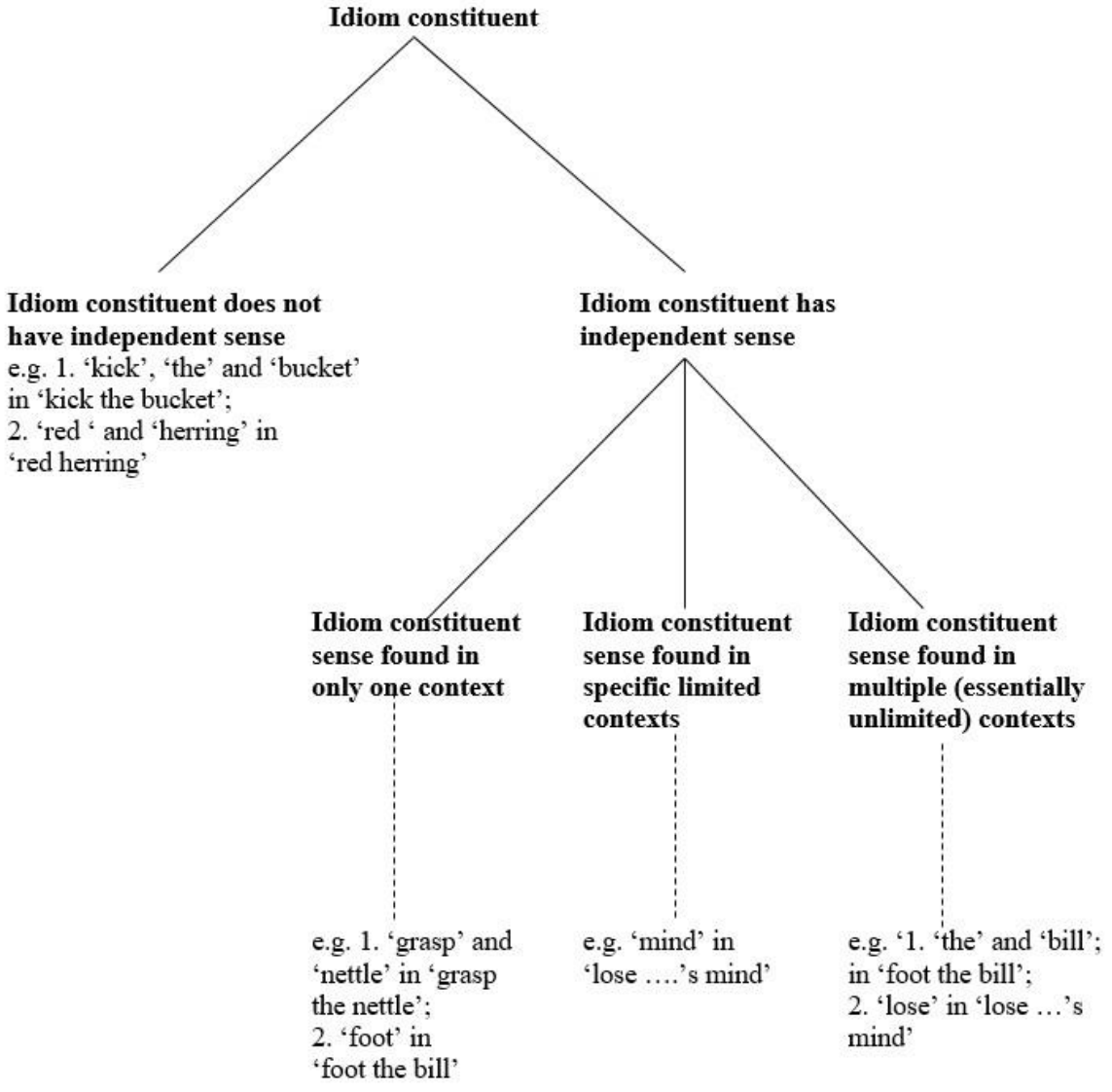

Figure 16: Analysis of idioms according to semantic independence of constituents

As with other multiword expressions, constituents in idioms may not have an independent sense, i.e. the constituent cannot be analysed semantically independently from other constituents in the idiom. Alternatively, the constituent may have an independent sense, i.e. it can be analysed semantically independently from other constituents in the idiom. Where a constituent has an independent sense, this may occur either (i) only in the context of this idiom,

(ii) in a limited number of other contexts as well, or (iii) in multiple contexts (essentially unlimited in number).

Idioms involving constituents which do not have independent senses must have logically at least two such constituents – these two constituents (neither with an independent sense) together forming a larger constituent which does have independent sense. Thus the entire idiom 'kick the bucket' has an independent sense (= die), but none of the constituents which make it up (taking each word in the idiom to be a separate constituent), 'kick', 'the' or 'bucket', has an independent sense. Where idioms are made up entirely of constituents which do not have independent senses, these idioms cannot be changed in any way (although 'additions' may be made, e.g. for tense, as in 'he kicked the bucket'). So, for example, one cannot say 'the bucket was kicked', or 'They both kicked buckets' (cf. Section 4).

Another example of idiom constituents which do not have an independent sense are 'red' and 'herring', in 'red herring', i.e. 'anything that diverts attention from a topic or line of inquiry' (CEDO). 'Red herring' is interesting because although it is typically classified as an idiom, it could arguably also be classified as a compound (for the analysis of compounds, see sections 5–5.3).

A second type of idiom constituent has an independent sense, but this sense is only found in the context of this idiom. An example is 'grasp' in 'grasp the nettle'. Here, 'grasp' has the sense 'tackle', 'deal with' or similar. This is not, however, a sense which 'grasp' has in any other context. Similarly, 'nettle' in 'grasp the nettle', has the sense 'difficult problem' or similar, a sense which it does not have in any other context. The fact that these two constituents (words) have independent senses in the idiom 'grasp the nettle' is shown by the fact that the constituents in the idiom (unlike those in 'kick the bucket') can be reorganised grammatically and be further modified. For example, it is possible to say things like 'That's one nettle which you are just going to have to grasp', or 'Eventually the British government grasped the nettle of Irish peace' (cf. also Dickins 1998: 241–243, 324, 435).

There are also idioms in which only one element has an independent sense, this sense being only found in the context of this idiom. An example is 'foot' in 'foot the bill' (= 'pay or settle (a bill, esp. one which is large or unreasonable, or which has been run up by another party'). Here, 'bill' means what it does in unlimited other contexts. 'Foot', however, is only found with this sense in this one phrase[14].

A third type of idiom constituent is one which has an independent sense this being found in more than one context (distinguishing it from the second type of idiom constituent above), but only in a limited number of contexts. An example is 'mind', meaning roughly 'rational faculties', in 'lose …'s mind'. This seems to occur in only one other context: 'out of …'s mind'. That 'mind' has an independent sense here is shown by the demonstration that 'lose' has an independent sense – in fact a sense which occurs in unlimited contexts, e.g. 'lose ….'s sanity/rational faculties/self-control/temper'. It is noteworthy, also, that while it is possible

[14] It could be argued that there is marginally one other context in which 'foot' occurs in this sense. This is with 'it', for example in 'I can't pay this bill. You're going to have to foot it'. As this example shows, however, not only does 'bill' have to be somewhere in the general discourse-context; the word 'it' also has to refer specifically to 'bill'. It is, of course, also possible to say things like, 'I've been footing all the bills for months now', with 'bills' in the plural. Here, however, it seems sensible to say that 'foot' still occurs in the context of 'bill', albeit that 'bill' itself is in the plural form (i.e. it has the plural suffix 's').

to say 'regain ….'s sanity/rational faculties/self-control/temper', it is not possible to say *'regain …'s mind'.

A final type of idiom constituent is one which has an independent sense and this is found in multiple (essentially unlimited) contexts. Examples are 'the' and 'bill' in 'foot the bill', and 'lose' in 'lose …'s mind' (discussed immediately above). Nunberg, Sag and Wasow (1994) refer to idioms in which none of the constituents has an independent sense as 'idiomatic phrases', while Sag et al. (2002) refer to them as 'non-decomposable idioms'. Nunberg, Sag and Wasow (1994) refer to idioms in which each of the elements has an independent sense as 'idiomatically combining expressions' or 'idiomatic combinations', while Sag et al. (2002) refer to them as 'decomposable idioms'.[15]

Figure 17 shows the semantic relationships between collocations, formulaic sequences (formulaic language), multiword expressions, compounds, phrasal verbs and idioms.

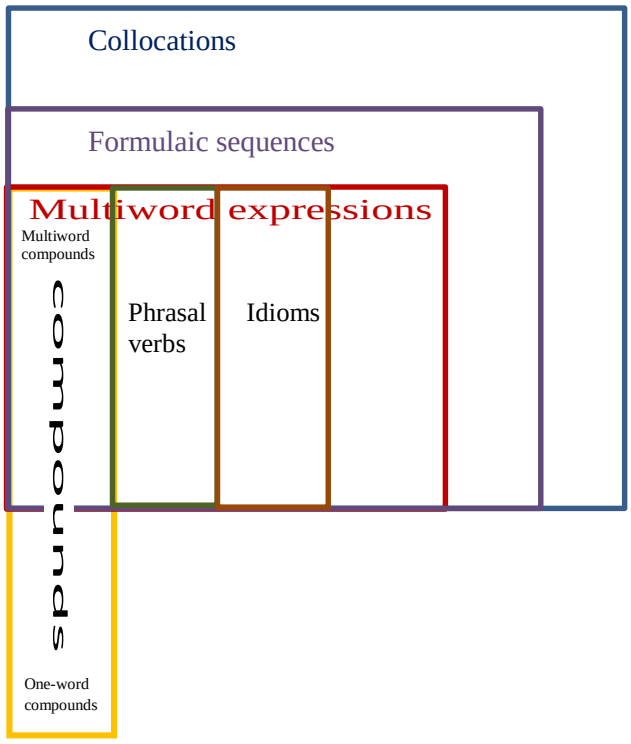

Figure 17: Semantic relationship between collocations, formulaic sequences (formulaic language), multiword expressions, compounds, phrasal verbs and idioms

In this section, I have defined an idiom as 'a phrase that is not a compound, not a phrasal verb, is non-clausal, and is not fully free-compositional'. I have also suggested that 'idiom' is a Group 1 term (Section 1), i.e. a non-technical everyday term. Given that 'compound' (sections 5–5.3) and 'phrasal verb' are semi-technical terms (Group 2 terms; Section 1),

[15] As with phrasal verbs, the account of the semantics of idioms which I have given here is not complete (see also Section 1). In addition to the denotative semantic independence of the components of idioms, there are also connotative issues of 'reflected meaning' (e.g. Dickins, Hervey and Higgins 2017: 103–104), and, associated with this, metaphor (cf. Dickins, Hervey and Higgins 2017: 194–210, and, rather more rigorously, Dickins 2005; Dickins 2018).

which we have some liberty in defining for our purposes, there does not seem a problem in defining these two notions in such a way that they are fully distinct (disjunct, in set-theoretical terms).

The greater problem is between i. compounds and idioms; and ii. phrasal verbs and idioms. Given that 'idiom' is a Group 1 term, a non-technical everyday term (Section 1), we are not in practice at much liberty to redefine it for academic purposes. The kind of issues this raises can be seen in relation to 'sleeping policeman', analysed in Section 5.1 as a multiword compound, and 'red herring', analysed in this section as an idiom. Both might be thought of as figurative (though this is not, in the current approach, a defining feature of either compounds or idioms): 'sleeping policeman' is fairly clearly metaphorical, while 'red herring' is, from a figurative perspective, puzzling, because there is no clear reason why a 'red herring' ('something that diverts attention from a topic or line of inquiry') should be figuratively related to a red-coloured herring (fish). It is also not immediately clear why 'sleeping policeman' should be analysed as a compound and 'red herring' as an idiom.

A partial grammatical solution could be found by further defining compounds as being members of specific word-classes: nominal, adjectival, adverbial, verbal, etc. This would make it possible to unambiguously categorise 'grasp the nettle' as an idiom, rather than a compound, because it involves more than one word class, having one constituent 'grasp', which is verbal, and another '(the) nettle', which is nominal. This does not help in the case of 'sleeping policeman' and 'red herring', however; both are nominal.

One way round this would be to rely on native speaker judgements, assuming fairly consistent judgements are made by native speakers in regard to idioms and non-idioms. Thus, we could define a compound as any multiword expression which is not regarded by native speakers as an idiom or a phrasal verb. Thus, if native speakers typically regard both 'sleeping policeman' and 'red herring' as idioms, we would class these both as idioms. If they regard only 'red herring' as an idiom, we would then regard (classify) 'sleeping policeman' as a compound (to be further discussed below).

We could adopt the same procedure with regard to phrasal verbs and idioms, i.e. taking native speakers' views on the nature of what is and is not an idiom into account first, and then classifying as a phrasal verb any relevant example which is not generally considered by them to be an idiom. Under this approach, if native speakers were to regard 'turn on' as an idiom, but 'give up' as not an idiom, the former would be classified as an idiom and the latter as a phrasal verb. This approach, however, seems rather messy, since 'phrasal verb' is fairly well understood as a semi-technical term, and what we have identified as phrasal verbs are perhaps only marginally regarded in everyday understanding of idioms as also being idioms. It may be better, then, on this basis to treat 'phrasal verb' and 'idiom' as discrete sets.

The distinction between compounds and idioms (e.g. putatively 'sleeping policeman' and 'red herring'), by contrast, seems much more problematic. The relative technicality of 'sleeping policeman' might make us somewhat more inclined to regard it as compound rather than an idiom. However, we can see that there are likely to be a very large number of similar examples in which a division such as technical vs. non-technical (etc.) is very blurred. An alternative to the definition of 'idiom' given at the start of this section and visualised as in Figure 17, therefore, might be to define idiom and compound as overlapping, i.e. to accept that a phrase can be both a compound and an idiom. We could, of course, also do the same

with ‘phrasal noun’ and ‘idiom’, making them overlapping classes, such that a phrase can be both a phrasal noun an idiom. These possible redefinitions are represented in Figure 18.

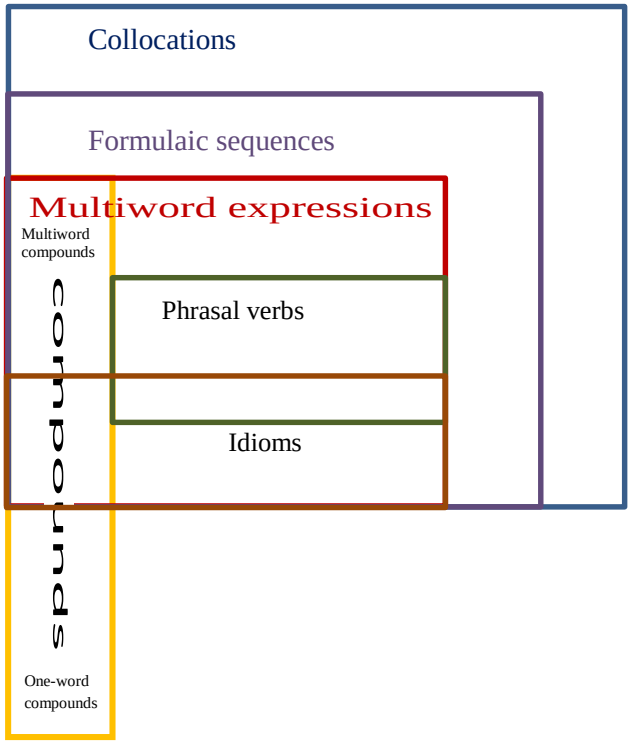

Figure 18: Semantic relationship between collocations, formulaic sequences, multiword expressions, compounds, phrasal verbs and idioms – with compounds and idioms, and phrasal verbs and idioms defined as overlapping classes

In the remainder of this article, I will stick with the definitions of compound, phrasal verb and idiom given at the start of sections 5, 6 and 7.

8. Proverbs

Unlike idioms, proverbs are clausal, i.e. they can stand on their own as sentence[16] (though they may also occur as a clause within a larger sentence, e.g. ‘Too many cooks spoil the broth, as they say’). OEDO defines a proverb as ‘A short, traditional, and pithy saying; a concise sentence, typically metaphorical or alliterative in form, stating a general truth or piece of advice; an adage or maxim’. Issa (2014, Chapter 2) provides a useful survey of different definitions of proverbs, as well as views on their typical (though not necessarily defining) features.[17]

[16] Not all proverbs have the standard form of sentences in English, and in particular they do not all contain a main verb, for example ‘any port in a storm’. All proverbs, however, can stand on their own as sentences, i.e. they can appear as complete utterances.

[17] Reviewer 1 has pointed out to me that there is an interesting difference between the definition of ‘proverb’ given here and the definitions of other categories discussed in this article (collocations, formulaic sequences,

Since 'proverb' is a non-technical term (Section 1), I will adopt the view here that proverbs are what native speakers consider to be proverbs (apart from the stipulation, above, that proverbs are clauses – which would, I believe, be reflected in terms of their identification of proverbs in practice by native speakers). This, of course, means that the boundaries between what is and is not a proverb will be somewhat fuzzy (Section 12): we cannot expect all native speakers to recognise exactly the same things as proverbs (and non-proverbs).

Some proverbs are fully free-compositional – i.e. all the words which make them up are used in the same sense in which they are used in other contexts. Examples are 'Honesty is the best policy', 'A little learning is a dangerous thing' and 'Better late than never'. Most proverbs, however, contain at least some words which are not completely free-compositional, i.e. they include words which are not found in the same sense in unlimited contexts. Examples are 'Too many cooks spoil the broth', 'A stitch in time saves nine', and 'Birds of a feather flock together' (all of which are fully bound-compositional; i.e. none of the constituent words in them has an independent sense). Fully free-compositional proverbs do not belong to the multiword expression category, though they do belong to the category of formulaic sequences (given the inevitably high levels of collocation they involve). Proverbs which contain at least some words which are not completely free-compositional belong to the multiword expression category.

In terms of the semantic independence of their constituents, proverbs can be analysed as illustrated in Figure 19 (next page).

'One swallow doesn't make a summer' is an example of a proverb in which none of the individual constituent words has an independent sense (the proverb's meaning being along the lines 'It should not be assumed that something is true just because there is one piece of evidence for it').[18] 'Silence is golden', meaning 'Silence is virtuous/preferable (to speaking)/to be enjoyed, etc.', provides an example of a second type of proverb constituent. Here 'golden' has an independent sense, but this sense is only found in the context of this proverb.[19] I have not been able to find an example of a third type of proverb constituent, where this constituent has a sense which only occurs in specific limited contexts.

The final type of proverb constituent, which has an independent sense and this sense is found in multiple (essentially unlimited) contexts, is illustrated by the proverb 'honesty is the best policy'. Here all the constituent words have the same sense as they have in unlimited

multiword expressions, compounds, phrasal verbs and idioms). In the case if proverbs alone, situational/discoursal considerations are invoked, e.g. in the OEDO definition 'A short, traditional, and pithy saying [...] stating a general truth or piece of advice […]'. There is thus inconsistency in the criteria used for identifying proverbs as opposed to these other categories. However, given that 'proverb' is a non-technical term (Section 1) – and that proverbs can, I think, can only be reasonably identified on a non-technical basis – the characterisation of proverbs in situational/discoursal terms is justified.

[18] I believe that the analysis that none of the constituents in 'one swallow doesn't make a summer' has an independent sense is correct. However, as Reviewer 2 has pointed out to me, this might be queried. He/she notes, "One wonders whether this is really true for *one* (and perhaps also for the negation). After all, the author invokes the idea of oneness in the characterization of the proverb's meaning ("It should not be assumed that something is true just because there is **one** piece of evidence for it")."

[19] Other similar senses of 'golden', such as 'Of time, an opportunity: Of inestimable value; exceedingly favourable or propitious' (OEDO), as in 'a golden opportunity' are, I believe, distinct from the sense of 'golden' in 'silence is golden', having a clearly different semantic range.

other contexts, i.e. the proverb is fully free-compositional. Another example is provided by 'silence' and 'is' in 'silence is golden'.[20]

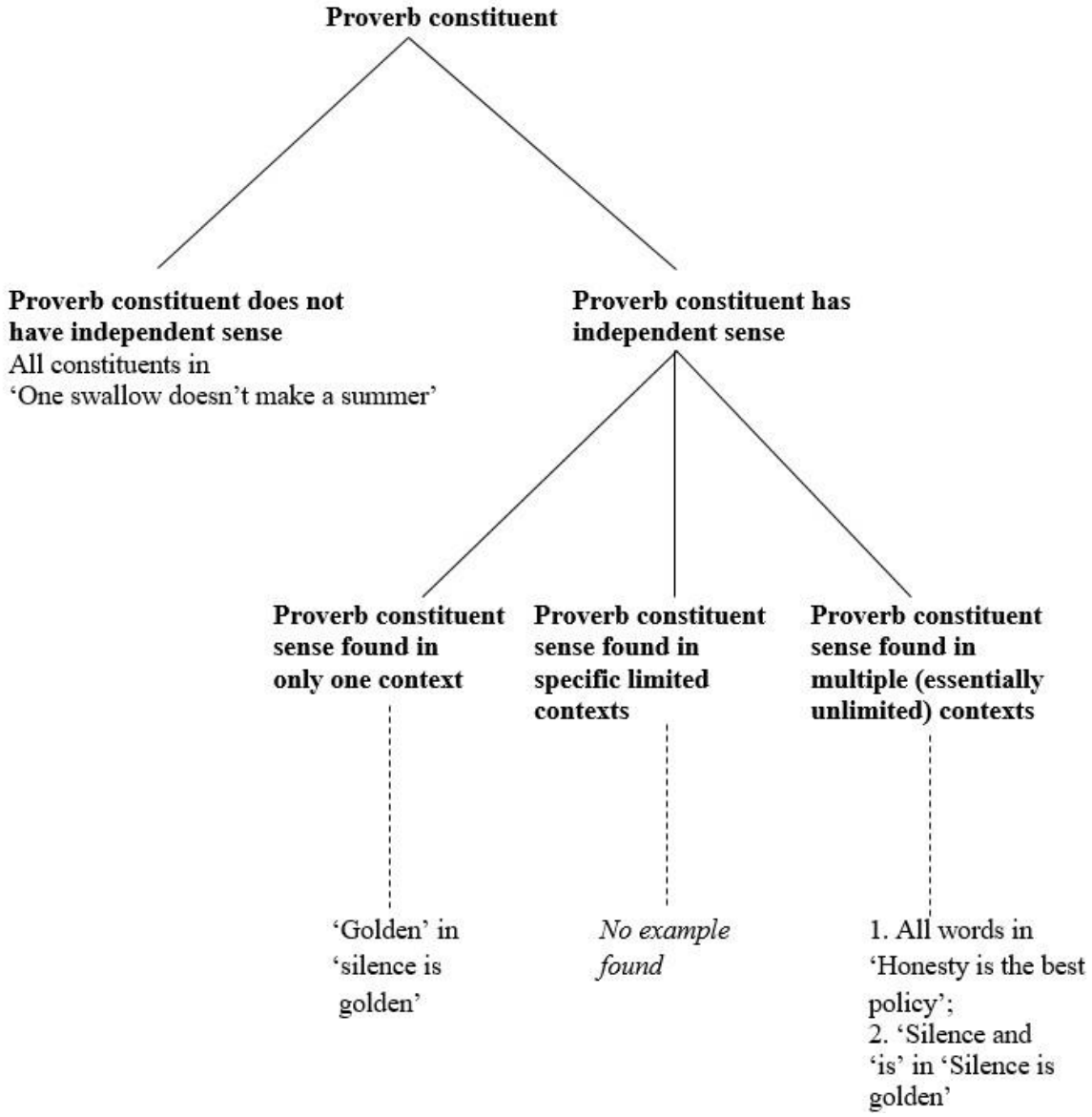

Figure 19: Analysis of proverbs according to semantic independence of constituents

Figure 20 (next page) shows the semantic relationship between collocations, formulaic sequences (formulaic language), multiword expressions, compounds, phrasal verbs, idioms and proverbs.

[20] As with phrasal verbs and idioms, the account of the semantics of proverbs which I have given here is not complete (see also Section 1). In addition to the denotative semantic independence of the components of proverbs, there are also connotative issues of 'reflected meaning' (e.g. Dickins, Hervey and Higgins 2017: 103–104), and, associated with this, metaphor (cf. Dickins, Hervey and Higgins 2017: 194–210, and, rather more rigorously, Dickins 2005; Dickins 2018).

9. Other types of multiword expression?

In this article, I have assumed that there are no other types of multiword expression (as defined in this article: Section 4) in English in addition to compounds, phrasal verbs, idioms and some proverbs[21]. This would need to be tested in future research. There are, of course, other terms for phenomena which are deemed to be similar to proverbs, such as 'aphorism', 'maxim', 'saying' and 'adage'. These typically belong under the category of 'formulaic sequence' and are not multiword expressions.

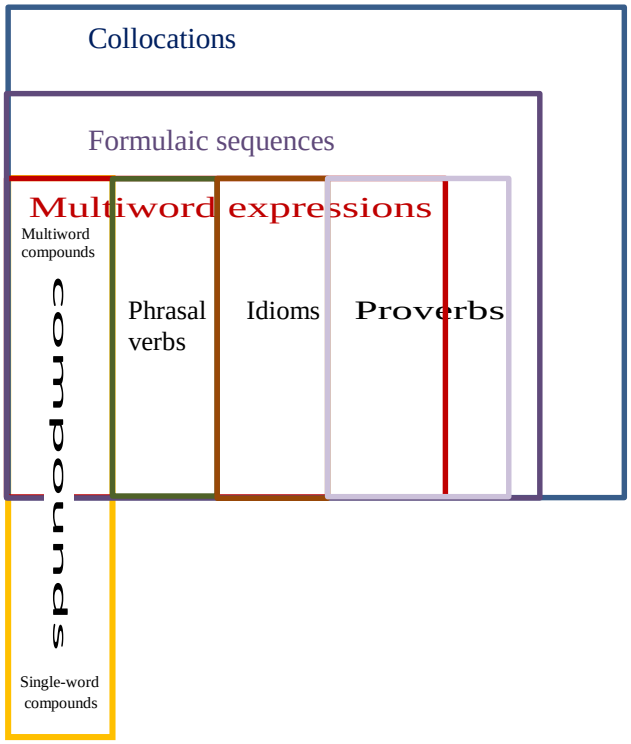

Figure 20: Semantic relationship between collocations, formulaic sequences, multiword expressions, compounds, phrasal verbs, idioms and proverbs

10. Further categories deriving from 'collocation', 'formulaic sequence' and 'multiword expression'

I have established the basic categories of collocation, formulaic sequence and multiword expression. I have also established the more specific categories of phrasal verb and idiom overlapping with multiword expression, and of proverb overlapping with multiword expression and non-MWE formulaic sequence (see below).

[21] I have excluded proper names from consideration in this article; these are, for example, included as a type of multiword expression by Sag, Baldwin, Bond, Copestake and Flickinger (2002) in their discussion of multiword expressions. They have, however, specific features which I believe mean that they need to be analysed separately from the phenomena considered in this article.

On the basis of the basic categories of collocation, formulaic sequence and multiword expression we identify further analytically useful categories, as follows:

1. **Non-formulaic collocation**; i.e. a collocation which is not a formulaic sequence (formulaic sequences of course also include multiword expressions, which themselves properly include multiword compounds, phrasal verbs, idioms, and proverbs) (see Section 10).
2. **Non-MWE formulaic sequence** (where 'MWE', as noted in Section 4, stands for 'multiword expression'); i.e. a formulaic sequence which is not a multiword expression (multiword expressions also include multiword compounds, phrasal verbs, idioms, and fully free-compositional proverbs).
3. **Non-MWE collocation**; i.e. a collocation which is not a multiword expression (multiword expressions also include multiword compounds, phrasal verbs, idioms, and fully free-compositional proverbs).

Figure 21 shows the scope of category 1 *non-formulaic collocations*: this is the area with the green background.

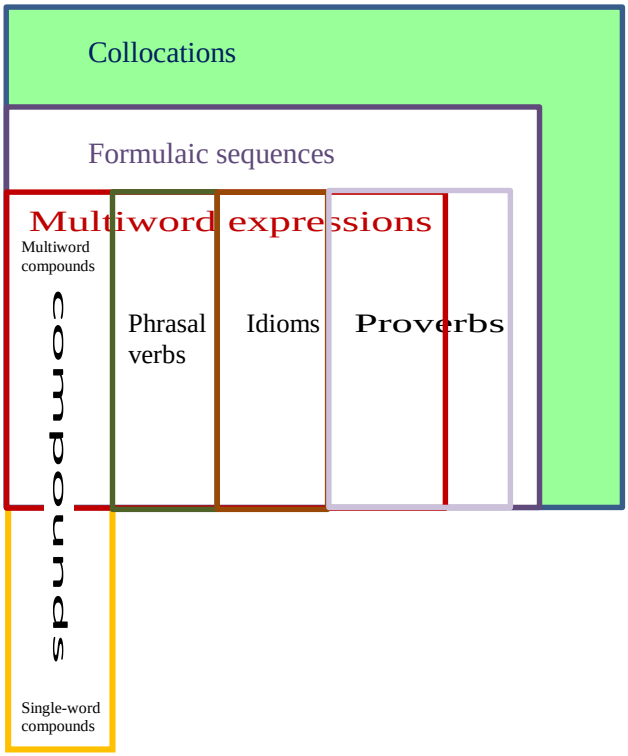

Figure 21: Scope of non-formulaic collocations

Category 1 (Figure 21), *non-formulaic collocations*, comprises fully free-compositional collocations which, however, lack the syntactic coherence (and possibly also the statistical frequency; see Section 3) to count as formulaic sequences. Because of their lack of syntactic coherence (and possibly the fact that the statistical occurrence of the collocational elements is not very significantly higher than their statistical occurrence across all contexts), these are the kind of collocations which may not be recognised by native speakers as such, and are likely to be only revealed by statistical computational analysis.

Figure 22 shows the scope of non-MWE formulaic sequences – the area with the green background.

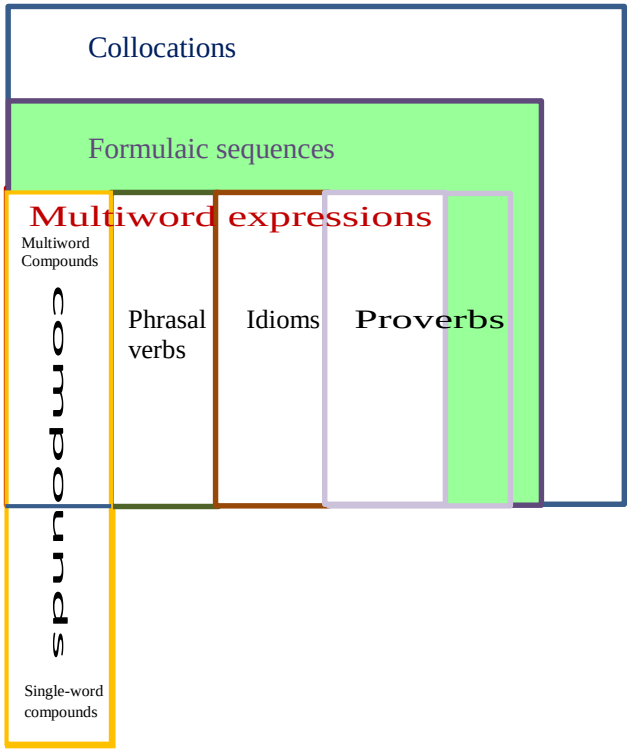

Figure 22: Scope of non-MWE formulaic sequences

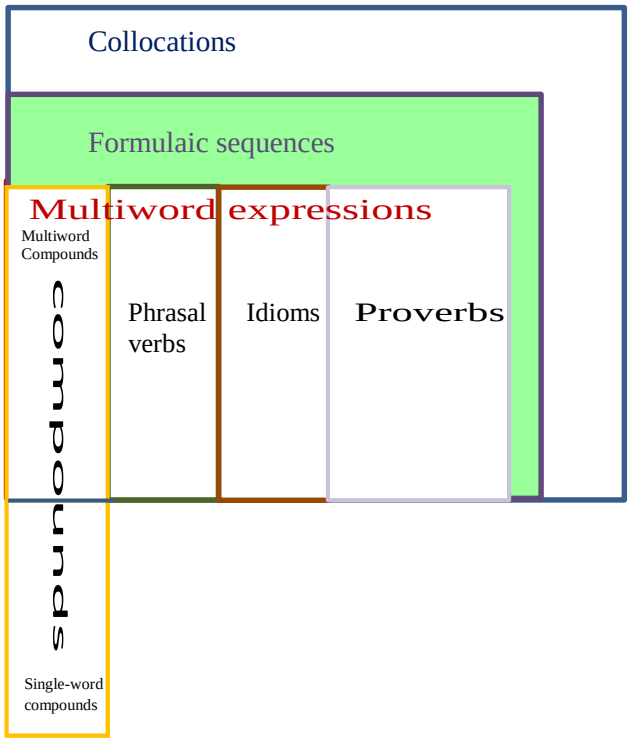

Figure 23: Scope of non-MWE, non-proverbial formulaic sequences

Category 2 (Figure 22), *non-MWE formulaic sequences*, comprises fully free-compositional collocations whose syntactic coherence (and possibly also frequency of occurrence) means they are classified as formulaic. One potential practical issue with category 2, *non-MWE formulaic sequences*, is that it cuts through proverbs: those proverbs which are fully free-compositional are included in it, while those that are not (i.e. which are multiword expressions) are excluded from it. As noted (Section 1), 'proverb' is an everyday term, for which native speakers are likely to agree on the ostensible definition (i.e. they are likely to agree about are and are not cases of proverbs), even if they cannot provide an abstract definition of what a proverb is in principle. 'Proverb' is, accordingly, for native speakers a fairly natural class (for which see the discussion in Hervey 1982: 17–18, of Peirce 1960) and not one which one would want to split up by another taxonomy. Accordingly, instead of the category *non-MWE formulaic sequences*, it would probably be preferable to operate with an otherwise identical category which excludes all proverbs, i.e. *non-MWE, non-proverbial formulaic sequences*. This is represented by the green-shaded area in Figure 23 (previous page).

Figure 24 shows the scope of category 3 above, *non-MWE collocations*. This category is useful for distinguishing between those collocations and formulaic sequences which are fully free-compositional and those which are not.

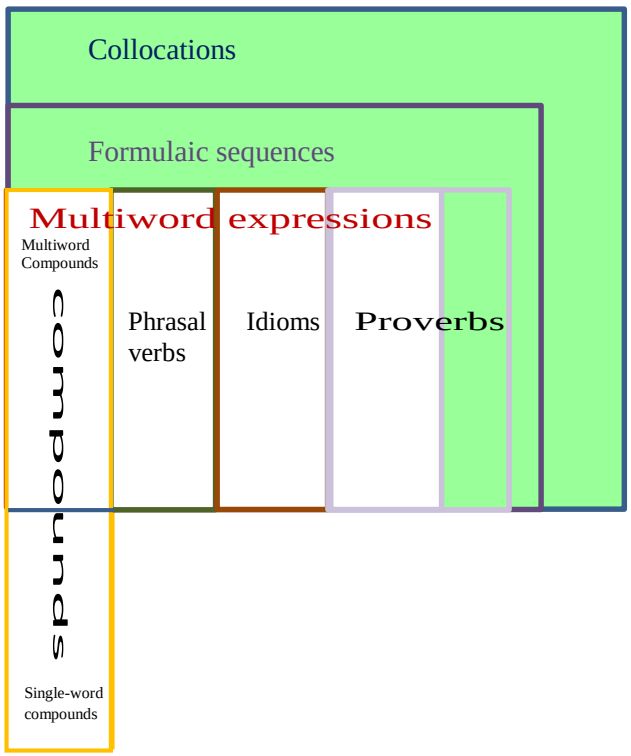

Figure 24: Scope of Non-MWE collocations

As with category 2, *non-MWE formulaic sequences*, category 3, *non-MWE collocations* seems unnatural because it cuts through proverbs, including some and excluding others. As with category 2, therefore, it seems better to operate with an otherwise identical category here which excludes all proverbs, i.e. *non-MWE, non-proverbial collocations*. This is represented by the green-shaded area in Figure 25 (next page).

In the approach known as ‘phraseology’, the term ‘collocation’ is sometimes defined to mean roughly what is meant in this article by ‘non-MWE collocations’, such that collocations “differ from other types of phraseological units that exhibit a fixed form and a non-decomposable, unitary meaning” (Pastor 2017: 29). For an introduction to phraseology, see Gries (2008: 3–26). Cruse (1986: 40) also defines ‘collocation’ in this way.

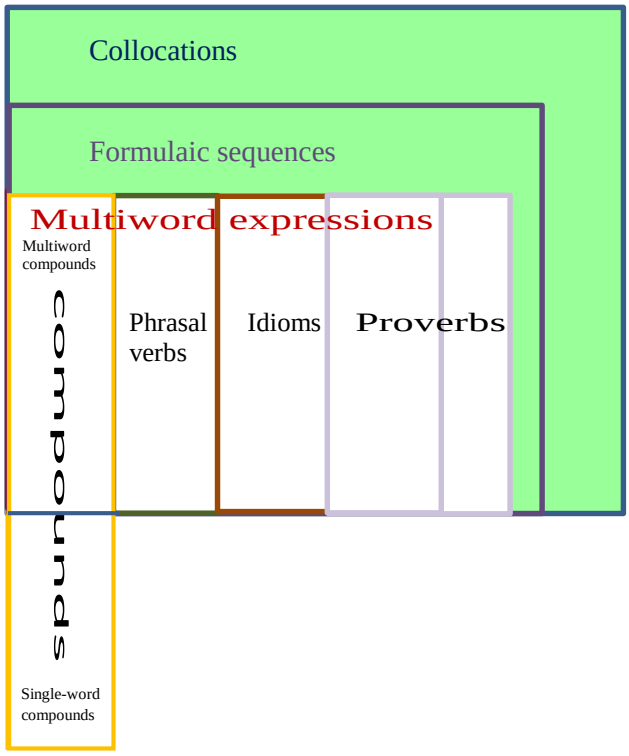

Figure 25: Scope of Non-MWE, non-proverbial collocations

The category of non-MWE, non-proverbial collocations allows for the investigation of all those fully free-compositional collocations, whether weak or strong (excluding proverbs), ignoring the question (which may be a moot one) of whether the collocations in question constitute formulaic sequences. Other derived categories in addition to those identified in this section could, of course, be established, depending on specific research concerns.

11. Semantic correlates of syntactic relationships in multiword expressions

Up till now I have focused on constituents, and largely ignored the semantic correlates of syntactic relationships. It is clear that these semantic correlates, even in fully free-compositional expressions are varied and can be complex (e.g. Ó Séaghdha 2008). For instance, a ‘door key’ is normally a key for opening a door, while an ‘ignition key’ is normally a key for turning on the ignition. It is, however, perfectly possible to say ‘door and ignition keys’, demonstrating that ‘key’ has the same sense in both ‘door key’ and ‘ignition key’, a sense which is also found essentially unlimited other contexts. It is also clear that ‘door’ and ‘ignition’ in ‘door key’ and ‘ignition key’ have the same sense as they do in unlimited other contexts, as can be seen from the possibility of phrases such as ‘keys for the door, the ignition,

the garage, the house, the garden shed [etc.]'. English (other languages also) allows for semantic narrowing in given phrases of the full possible denotative range which can be expressed by, for example, noun-noun expressions. It falls outside the scope of this article to investigate this in further detail, but it is an issue which needs to be addressed for a full understanding of the compositionality or non-compositionality of collocations. It may be that a distinction needs to be made between forms like 'door key' and 'traffic lights'. 'Door key' arguably involves an extra-linguistically determined (or, perhaps better, 'motivated') restriction on typical range of reference – what Dickins, Hervey and Higgins refer to as associative connotative meaning (Dickins, Hervey and Higgins 2017: 97–99; for a more theoretically grounded account, see Dickins 2014). 'Traffic lights', by contrast, seems to have a sense which is much more specifically denotatively fixed (cf. Sag, Baldwin, Bond, Copestake, and Flickinger 2002). In 'traffic lights', just as much as 'door key', however, the two constituent words both seem to have an independent sense. This is borne out by the fact that it is possible to say things like 'traffic lights and signals' (15 results on IWeb, 1.10.18: <https://corpus.byu.edu/iweb/>), where at least some of the examples are to be read as equivalent to 'traffic lights and traffic signals'; e.g. "Ensuring that traffic flows as smoothly as possible at all times, including rush-hours, calls for planning, studying traffic volumes at various times of day, and knowing how to install and coordinate traffic lights and signals, signs and other traffic control devices": <https://www.cityoftulsa.org/government/departments/streets-and-stormwater/streets/>. It is also borne out by the fact that one can also say things like 'street and traffic lights' (21 results on IWeb, 1.10.18: <https://corpus.byu.edu/iweb/>) or 'traffic and street lights' (4 results on IWeb, 1.10.18: <https://corpus.byu.edu/iweb/>), where at least some of the examples are to be read as equivalent to 'traffic lights and street lights'; e.g. "The road system and infrastructure would likely need major upgrades for driverless vehicles to operate on them. Traffic and street lights, for instance, would likely all need altering": <https://axleaddict.com/safety/Advantages-and-Disadvantages-of-Driverless-Cars>.

It is possible that what differentiates 'traffic lights' from 'door key' semantically is the greater specificity of the denotative relationship associated with the noun-noun syntactic relationship in the former as compared to the latter. It is, finally, worth stressing, regardless of issues of the semantic correlates of syntactic relationships between elements in phrases of this kind, that these correlates fall into definable patterns. This is evidenced by the oddity of a neologism such as 'tamper-evident' as in 'tamper-evident packaging' (i.e. packaging where it is/will be evident if it has been tampered with). Here, the words 'tamper' and 'evident' have their standard senses, but the semantic correlates of the syntactic relationship involved falls outside the standard range permitted by the English language.

12. Fuzzy boundaries and discrete boundaries

In understanding categories, it is essential to distinguish between those which involve discrete boundaries and those which involve fuzzy boundaries. Fuzzy boundaries involve situations in which it is not entirely clear which of two classes a particular entity is to be assigned to. This situation is to be distinguished from semantic overlap, where a particular entity can be assigned to two classes simultaneously. Semantic overlap is illustrated by 'doctor' and

'genius'. Some, but not all, doctors are geniuses, and some, but not all, geniuses are doctors. We can accordingly say 'he is both a doctor and a genius' (or 'he is both a genius and a doctor'). This situation contrasts with, for example, 'cup' and 'mug'. Labov (1972) showed that English speakers distinguish cups from mugs according to a variety of features including the shape of the vessel concerned. There are also numerous objects which a speaker might describe as a cup or a mug (and one might even say, for example, 'You could call that a cup or a mug'). However, 'cup' and 'mug' are not a case of semantic overlap: one cannot say 'that is both a cup and a mug' (or 'that is both a mug and a cup'). 'Cup' and 'mug' are, in abstract semantic terms, discrete (in set-theory 'disjunct') notions. In the real world, however, the boundary between 'cup' and 'mug' is fuzzy. (Cf. Dickins 2014: 20, for further discussion of abstract semantic disjunction vs. real-world (realisational) 'semantic overlap', i.e. fuzzy boundaries.)

We have already seen that in perhaps the most basic distinctions made in this article – those between everyday terms, semi-technical terms and technical terms (Section 1) – we are dealing with fuzzy boundaries. It will not in all cases be clear whether a particular term should be regarded as everyday or semi-technical, or as semi-technical or technical. We can consider in the light of the same distinction between fuzzy boundaries and discrete boundaries the fundamental concepts defined in this article: collocations, formulaic sequences, multiword expressions, compounds, phrasal verbs, idioms and proverbs.

At the outer boundary of these concepts is the distinction between collocation and non-collocation. If we use statistical frequency as our sole criterion, this is, properly speaking, a discrete boundary. Technically, as soon as we have a frequency of co-occurrence between two or more words which is higher than predicted by their frequency of occurrence as individual words, we have a collocation. In practice, we are unlikely to choose to investigate collocations where this frequency of co-occurrence is not significantly greater than the frequency of occurrence of the individual words; but this is a matter of research focus, not of the definition of what constitutes a collocation.

The distinction between non-formulaic collocation and formulaic sequence, by contrast, is potentially complex. If we use syntactic coherence as our sole criterion, we have a discrete boundary, as we do if we also add to this statistical frequency. The statistical frequency boundary is, however, conventional, because we could have chosen to draw it somewhere else. Finally, as soon as we introduce criteria involving native-speaker judgements, the boundary becomes fuzzy (with different native speakers no doubt making different judgements in different cases).

Since non-MWE collocations involve only fully free-compositional constituents and multiword expressions (MWEs) have at least one constituent which is not fully free-compositional, the boundary between the two is non-fuzzy – at least assuming that it is unambiguously possible to determine what is a fully free-compositional constituent and what is not.

The relationship between multiword expression and multiword compound, multiword expression and phrasal verb, and multiword expression and idiom are clear: all of multiword expression, multiword compound and idiom are sub-types (properly included in) multiword expression. There are no complications between multiword expression and these properly included concepts in terms of fuzzy boundaries.

The boundaries between compounds, phrasal verbs and idioms are more problematic. It seems possible to define compounds and phrasal verbs in a way which both makes the two

notions disjunct (in set-theoretical terms) and the boundary between them non-fuzzy, and does not noticeably conflict with the general understanding of these terms (cf. Section 7). The boundary between compounds and idioms can be made disjunct, though that boundary is likely to be fuzzy: even if we establish two disjunct classes, thereby stipulating that an example can only be a member of one class, there are going to be cases where we are not sure which class we wish to assign that example to. If, of course, we were to define 'compound' and 'idiom' as overlapping classes (Section 7), some examples would, by definition, potentially be members of both classes.

Similarly even if the boundary between phrasal verbs and idioms is made disjunct, this boundary is likely to be fuzzy, with examples in which it is not entirely clear what is the more appropriate category. If we were to define 'phrasal verb' and 'idiom' as overlapping classes (Section 7), some examples would, by definition, potentially be members of both classes. Finally, the boundary between compounds/phrasal verbs/idioms on the one hand, and proverbs on the other is non-fuzzy: only proverbs are clausal, and there should, therefore, be no indeterminate cases. 'Proverb', of course, overlaps with multiword expression and non-MWE formulaic sequence, meaning that a particular proverb may be also a multiword expression or a non-MWE formulaic sequence.

13. Universal categories and language-specific categories

The following categories are universal in that we would expect to find them in any language: collocation, formulaic sequence, multiword expression. For a natural language not to have collocations would involve the vanishingly unlikely situation that there was no variation in the statistical frequency of words regardless of what other words they occurred in the context of. Corresponding arguments apply to formulaic sequences. 'Multiword expression', as defined in this article, is a notion derived from the fundamental possibilities of the relationships between syntax and semantics. It is possible for 'languages' not to have multiword expressions (logical languages do not have them, for example). All natural languages, however, seem in practice to have multiword expressions.

The following categories are better regarded as non-universal: compound, phrasal verb, idiom and proverb. 'Compound' is a semi-technical term (a Group 2 term; Section 1), which is applied to specific lexical features in English. However, this may not be an appropriate term to use with other languages, for instance because their lexical features are very different from those of English. Alternatively, if we use the same term, we need to be aware that what we are referring to are not necessarily the same types of features as in English. Either way, 'compound' is not a universal category, meaning that we cannot simply transpose the ontology given in this article from English to other languages. The same applies to 'phrasal verb', which is not a category found in many languages; and even in languages where a category exists which can reasonably be given this name, it may well be very different in key respects from 'phrasal verb' in English. 'Idiom', as an everyday non-technical (Group 1; Section 1) term, is, similarly, very unlikely to have simple correspondents in many other languages, and can therefore be said to not exist as a 'ready-made' concept in these languages.

Although sayings which we might identify as 'proverbs' may be found in all natural languages (cf. Issa 2014: Section 2.2), we should not take it that the category of 'proverb' as

defined in (and applied to) English is itself universal. Arabic, for example, makes a distinction between a *maṯal*, typically translated as 'proverb' and a *ḥikma* (literally 'wisdom'), i.e. a wise saying. Not all examples of what Arabs regard as *maṯal* seem to be what English speakers would think of a proverbs, while *ḥikma* is a category which does not really exist for English. The notion 'proverb' cannot therefore be simply transposed from this ontology for English to one for Arabic, or by extension to other languages.

14. Conclusion

The article has attempted to establish an ontology (with discussion of some alternatives) for collocations, formulaic sequences, multiword expressions, compounds, idioms and phrasal verbs in English. I have stressed the 'constructive' nature of this: particularly in the case of technical terms, it is not a question of saying 'what do these terms mean?', but rather: 'what can we use these terms to mean, such that we arrive at an overall ontology which is both coherent and useful for our purposes?'. I have also stressed that different kinds of word/phrases for notions in the ontology should, for practical reasons, be treated rather differently: we have more leeway to redefine technical terms for our own purposes (particularly when these terms already have multiple and contradictory definitions in the literature) than we do to redefine everyday terms which already have a fairly generally agreed (if probably rather vague) meaning. The resulting ontology is not intended to be definitive. We might expect it to be improved on through further consideration of the same areas of phenomena. It is also only of value for investigating the areas of phenomena which it seeks to define. For other, even closely related, areas of analysis, other ontologies involving some other terms and notions would need to be established.

References

BAHNS, Jens and Moira ELDAWa. 1993. 'Should we teach EFL students collocations?'. *System*, 21(1). Pp. 101–114.

BALDWIN, Timothy and Su Nam KIM. 2010. 'Multiword expressions'. In Nitin Indurkhya and Fred Damerau (eds.), *Handbook of natural language processing* (2nd ed.). Boca Raton: CRC Press. Pp. 267–292.

BARTSCH, Sabine. 2004. *Structural and functional properties of collocations in English: a corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen: Narr.

BAUER, Laurie. 2004. *A glossary of morphology*. Edinburgh: Edinburgh University Press.

CAI, Ying. 2017. *Second language acquisition of Chinese verb-noun collocations*. University of Massachusetts Amherst: Masters thesis.

CALZOLARI, Nicoletta, Charles FILLMORE, Ralph GRISHMAN, Nancy IDE, Alessandro LENCI, Catherine MACLEOD, and Antonio ZAMPOLLI. 2002. 'Towards best practice for multiword expressions in computational lexicons'. In *Proceedings of LREC 2002*. Canary Islands. Pp. 1934–1940.

CARPUAT, Marine and Mona DIAB. 2010. 'Task-based evaluation of multiword expressions: a pilot study in statistical machine translation'. In *Proceedings of NAACL/HLT 2010*. Los Angeles. Pp. 242–245.

CARSTAIRS-MCCARTHY, Andrew. 2002. *An introduction to English morphology*. Edinburgh: Edinburgh University Press.

CONSTANT, Mathieu, Gülşen ERYIĞIT, Johanna MONTI, Lonneke van der PLAS, Carlos RAMISCH, Michael ROSNER, and Amalia TODIRASCU. 2017. 'Multiword expression processing: a survey'. *Computational Linguistics*, 43(4). Pp. 837–892.

COWIE, Anthony. 1978. 'The making of dictionaries and reference works'. In Peter Strevens (ed.), *In honour of A.S. Hornby*. Oxford: Oxford University Press. Pp. 127–139.

CRUSE, David Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

CRYSTAL, David. 2008. *A dictionary of linguistics and phonetics* (6th edn.). Oxford: Blackwell.

DICKINS, James. 1998. *Extended axiomatic linguistics.* Berlin and New York: Mouton de Gruyter.

——. 2005. 'Two models for metaphor translation'. *Target* 17(2). Pp. 227–273.

——. 2006. 'The verb base in Central Urban Sudanese Arabic'. In Janet Watson and Lutz Edzard (eds.), *Arabic grammar as a window on Arab humanism: essays in honour of Michael Carter*. Wiesbaden: Otto Harrassowitz Verlag. Pp. 155–195.

——. 2014. 'Associative meaning and scalar implicature: a linguistic-semiotic account'. *Linguistica* ONLINE, <http://www.phil.muni.cz/linguistica/art/dickins/dic-003>.

——. 2018. 'Tropes and translation'. In Adelina Hild and Kirsten Malmkjaer (eds.), *The Routledge handbook of translation studies and linguistics*. London and New York: Routledge. Pp. 208–222.

DICKINS, James, Sándor G.J. HERVEY and Ian HIGGINS. 2017. *Thinking Arabic translation* (2nd edn.). London and New York: Routledge.

EVERT, Stefan. 2007. 'Corpora and collocations: extended manuscript': <http://www.stefan-evert.de/PUB/Evert2007HSK_extended_manuscript.pdf>.

FIRTH, John Rupert. 1957. *Papers in linguistics 1934–1951*. London: Oxford University Press.

GIBBS, Raymond W. Jr. 2010, 'Idioms and formulaic language'. In Dirk Geeraerts and Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*. Oxford and New York: Oxford University Press. Pp. 697–725.

GÓMEZ, Pascual Cantos. 2009. 'Attempting to model sense division for word sense disambiguation'. In Maria Manuela Cruz-Cunha, Eva F. Oliveira, Antonio J. Tavares, and Luis G. Ferreira, (eds.), *Handbook of research on social dimensions of semantic technologies and web services*, Vol. 2. (2 Volumes). Hershey: IGI Global. Pp. 126–157.

GRANT, Lynn. 2003. *A corpus-based investigation of idiomatic multiword units*. University of Wellington: PhD thesis.

GRANT, Lynn and Laurie BAUER. 2004. 'Criteria for re-defining idioms: are we barking up the wrong tree?'. *Applied Linguistics*, 25(1). Pp. 38–61.

GRIES, Stefan Th. 2008. 'Phraseology and linguistic theory: a brief survey'. In Granger, Sylviane and Fanny Meunier (eds.), *Phraseology: an interdisciplinary perspective*. Amsterdam and Philadelphia: John Benjamins. Pp. 3–26.

GRIES, Stefan Th. 2013. '50-something years of work on collocations: What is or should be next ...'. *International Journal of Corpus Linguistics*, 18(1). Pp. 137–165.

HAUSMANN, Franz J. 1989. 'Le dictionnaire de collocation'. In Franz J. Hausmann, Herbert E. Wiegand, and Ladislav Zgusta (eds), *Wörterbücher, dictionaries, dictionaries: ein internationales Handbuch zur Lexikographie*. Berlin: de Gruyter. Pp. 1010–1019.

HERBST, Thomas. 1996. 'What are collocations: sandy beaches or false teeth?'. *English Studies*, 77(4). Pp. 379–393.

HERVEY, Sándor G. J. 1982. *Semiotic perspectives*. London: George Allen and Unwin.

HILL, Jimmie. 2000. 'Revising priorities: from grammatical failure to collocational success'. In Lewis, M. (ed.), *Teaching collocation*. Hove: Language Teaching Publications. Pp. 47–69.

ISSA, Huwaida Jaber. 2014. *Proverbs, modified proverbs and curses in two novels of the Syrian coast*. University of Leeds: PhD thesis.

KILGARRIFF, Adam. 1992. *Polysemy*. University of Sussex: PhD thesis.

KÖVECSES, Zoltan. 2010. *Metaphor* (2nd edition). Oxford: Oxford University Press.

LABOV, William. 1972. 'The boundaries of words and their meanings'. In Charles James N. Bailey and Roger W. Shuy (eds.), *New ways of analysing variation in English*. Washington DC: Georgetown University Press. Pp. 340–373.

LEHECKA, Tomas. 2015. 'Collocation and colligation'. In Jan-Ola Östman and Jef Verschueren (eds.), *Handbook of pragmatics online*. Amsterdam and Philadelphia: John Benjamins. <https://benjamins.com/online/hop/articles/col2>. Pp. 1–23.

LEWIS, Michael. (ed.). 2000. *Teaching collocations: further developments in the lexical approach*. Hove: Language Teaching Publications.

LIU, Dilin. 2008. *Idioms: description, comprehension, acquisition, and pedagogy*. London and New York: Routledge.

LYONS, John. 1991. *Natural language and universal grammar: essays in linguistic theory 1*. Cambridge: Cambridge University Press.

MERRILL, Michael. 1995. 'Putting "capitalism" in its place: a review of recent literature'. *The William and Mary Quarterly*, 52(2). Pp. 315–326.

MOLLIN, Sandra. 2014. *The (Ir)reversibility of English binomials: corpus, constraints, developments*. Philadelphia: John Benjamins.

MULDER, Jan W.F. and Paul RASTALL. 2005. *Ontological questions in linguistics*. Munich: Lincom.

NESSELHAUF, Nadja. 2005. *Collocations in a learner corpus*. Philadelphia: John Benjamins.

NUNBERG, Geoffrey, Ivan A. SAG, and Thomas WASOW. 1994. 'Idioms', *Language* 70(3). Pp. 491–538.

Ó SÉAGHDHA, Diarmuid. 2008. *Learning compound noun semantics (Technical report, number 735)*. Cambridge: Computer Laboratory, University of Cambridge. <https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-735.pdf>.

O'DONNELL, Matthew Brook, Ute RÖMER, and Nick C. ELLIS. 2013. 'The development of formulaic sequences in first and second language writing: investigating effects of frequency, association, and native norm'. *International Journal of Corpus Linguistics*, 18(1). Pp. 83–108.

PASTOR, Gloria Corpas. 2017. 'Collocational constructions in translated Spanish: what corpora reveal'. In: Ruslan Mitkov (ed.). *Computational and corpus-based phraseology*. Berlin: Springer. Pp. 29–40.

PEIRCE, Charles Sanders. 1960. *Collected papers of Charles Sanders Peirce* (vols. 1–8). Hartshorne, C. and Weiss, P. (eds.). Cambridge (Mass.): Harvard University Press.

POPPER, Karl. 1986 [1957]. *The poverty of historicism*. London and New York: Ark.

POSIO, Pekka. 2015. 'Subject pronoun usage in formulaic sequences. Evidence from Peninsular Spanish'. In Orozco, Rafael, Carvalho, Ana Maria & Shin, Naomi (eds.), *Subject pronoun expression in Spanish: a cross-dialectal perspective*. Washington: Georgetown University Press. Pp. 59–74.

RICHARDS, Jack C., John PLATT, and Heidi WEBER. 1985. *Longman dictionary of applied linguistics*. London: Longman.

RUIZ YEPES, Guadalupe. 2017. 'Hybrid Methods for the Extraction and Comparison of Multilingual Collocations in the Language for Specific Purposes of Marketing'. In *Proceedings of Computational and corpus-based phraseology: recent advances and interdisciplinary approaches*. London: EUROPHRAS. Pp. 11–18. <http://www.tradulex.com/varia/Europhras2017-II.pdf>.

SAG, Ivan A., Timothy BALDWIN, Francis BOND, Ann COPESTAKE and Dan FLICKINGER. 2002. 'Multiword expressions: a pain in the neck for NLP'. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002)*. Mexico City. Pp. 1–15.

SAUSSURE, Ferdinand de. 1959. *Course in general linguistics* (translated by Wade Buskin). New York: McGraw-Hill Paperbacks.

——. 1975 [1916]. *Cours de linguistique générale*. Payot: Paris.

——. 1983. *Course in general linguistics* (translated by Roy Harris). Duckworth: London.

SINCLAIR, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

WANG, Xiaofei. 2018. *Quotation and truth-conditional pragmatics*. London and New York: Routledge.

WULFF, Stefanie. 2008. *Rethinking idiomaticity: a usage-based approach*. Continuum: London and New York.

WRAY, Alison 2002. *Formulaic language and the lexicon.* Cambridge University Press: Cambridge.

WRAY, Alison and Michael PERKINS. 2000. 'The functions of formulaic language: an integrated model'. *Language and Communication*, 20(1). Pp. 1–28.

Linguistica ONLINE. Issue 23. Published: May 7, 2020
http://www.phil.muni.cz/linguistica/art/dickins/dic-006.pdf
ISSN 1801-5336

SYMBOLISATION FOR EXTENDED AXIOMATIC FUNCTIONALISM[*]1

James Dickins (University of Leeds, UK, J.Dickins@leeds.ac.uk)

Abstract: This article presents a set of symbols for the linguistic, and semiotic, theory of extended axiomatic functionalism. Section 2 provides a visual representation for one of the two components of the theory, the signum ontology, plus the ancillary areas of general phonetics and general semantics. Section 3 presents the proposed symbols for the signum ontology and ancillary areas. Sections 4-4.5 provide exemplification: Section 4.1 for general-phonetic notions, Section 4.2 for general-semantic notions, Section 4.3 for morphontic (non-meaning-related) only notions, Section 4.4 for semantic (meaning-related) only notions, and Section 4.5 for both morphontic and semantic notions. Sections 5-5.5 consider the principles adopted in drawing up these symbols, as follows: retention of existing linguistic symbols (Section 5.1), use of identical symbols for morphontics and semantics, and plain vs. italic text (Section 5.2), degree of abstraction (Section 5.3) – instantiation, i.e. direct model for speech events (level 1) (Section 5.3.1), immediate generalisation from speech events (level 2) (Section 5.3.2), secondary generalisation from speech events (level 3) (Section 5.3.3), and signum level (level 4) (Section 5.3.4), and degree of peripherality or centrality (Section 5.4).

Section 6 provides a representation of (i) the second of the two components of extended axiomatic functionalism, the system ontology, and (ii) the overall theory, comprising the signum ontology and the system ontology (plus ancillary areas of general semantics and general phonetics). Section 6.1 considers the proposed symbols for the system ontology. Section 6.2 discusses issues involved in choosing appropriate symbols for the system ontology. Section 7 considers ways in which the symbols for both the signum ontology and system ontology can be simplified. Section 8 compares the symbols proposed in this article with those proposed by Mulder for 'standard axiomatic functionalism'.

Key words: axiomatic functionalism, extended axiomatic functionalism, standard axiomatic functionalism, symbols, symbolisation, ontology, signum ontology, system ontology

[*] Previously unpublished. [Editor's note]

[1] I thank Barry Heselwood and two anonymous reviewers for *Linguistica* ONLINE for reading draft versions of this article and making very useful comments on it. These have considerably helped improve the final version. At various points in this final version of the article, I address comments made by the two *Linguistica* ONLINE reviewers on the earlier draft which they read, referring to them, where appropriate, as Reviewer 1 and Reviewer 2.

1. Introduction

This article proposes and justifies a proposed symbolisation – by which I simply mean a set of symbols – for notions and entities in extended axiomatic functionalism. Extended axiomatic functionalism comprises two components, i.e. two ontologies, where an ontology is a "set of entities presupposed by a theory" (Collins English Dictionary). These two components are i. the signum ontology, together with two ancillary areas, general phonetics and general semantics; and ii. the system ontology. I will deal with the signum ontology and the ancillary areas of general phonetics and general semantics, and the symbolisations for notions and entities within this, and then move on to the system ontology. At the end of the article, I will compare this symbolisation with that proposed by Mulder for the standard version of axiomatic functionalism.[2]

Both the extended version of axiomatic functionalism (extended axiomatic functionalism) presented here, and the standard version – 'standard axiomatic functionalism' developed by Mulder and Hervey – are general semiotic as well as linguistic theories (for formal statements of both theories, see Dickins 2009; and Mulder and Hervey 2009). In extended axiomatic functionalism, in talking about semiotics generally (i.e. non-linguistic semiotic systems), *cen-* is used instead of *phon-* (Dickins 2009; 11, Def. 0b), and *log-* is used instead of *lex-* (Dickins 2009; 11, Def. 0b), while *del-* is used for both linguistic and non-linguistic semiotic systems. For written languages (as opposed to spoken languages) as semiotic systems, *graph-* is used instead of *phon-* (Dickins 2009; 11, Def. 0c). The symbols presented in this article can be used equally for natural language (spoken and written) and non-linguistic semiotic systems.

All the notions expressed by the symbols proposed in this article can also be expressed in other, existing, ways in extended axiomatic functionalism. These other ways are, however, less concise than those proposed in this article, most having the character of definitions; i.e. they involve combinations – sometimes extremely complex – of more than one, already previously defined, symbol. The symbols proposed in this article, by contrast, are all fairly simple. They are intended to be used in linguistic descriptions, for the purposes of clarity and precision, allowing (and requiring) the writer to state precisely what kind of entity they are referring to, and communicating this to the reader. As Reviewer 2 has pointed out to me, symbolisations involving combinations of previously defined symbols are also definitions – for *i*Rd for 'phonete', *f*Rd or {*i*}Rd for 'allophone' and {*f*}Rd for phonogical form / figure (all in the 'phonologics' column in Figure 1).[3] Simple symbols with which these are equated, e.g. *p* in the formula *p* = {*f*}Rd (Figure 1) are symbols but not definitions. Reviewer 1 points out that speed in physics is symbolized as *v*, but defined as *d* / *t* (distance divided by time); i.e. the former provides only a graphic symbol, while the latter states its relation to other physical notions.

[2] Reviewer 1 has pointed out that this symbolisation is purely representational, i.e. there is a one-one conventional relation between entity and symbol; there are no axioms or rules of deduction, so it is not a calculus.

[3] In Dickins (1998: 134), I proposed 'phonotics' as a cover term for 'phonetics', 'allophonics' and 'phonologics' (cf. Dickins 2009: 35, Def. F3f); cf. also 'morphotics' (Dickins 1998: 134; Dickins 2009: 42, Def. F1b1a2), 'semotics' (Dickins 1998: 135; Dickins 2009: 43, Def. F1b2a), and 'delotics' (Dickins 1998: 135; Dickins 2009: 45, Def. F4.1).

2. Signum ontology

Figure 1 provides a visual representation of the signum ontology of extended axiomatic functionalism, as well as the ancillary areas of general phonetics and general semantics.

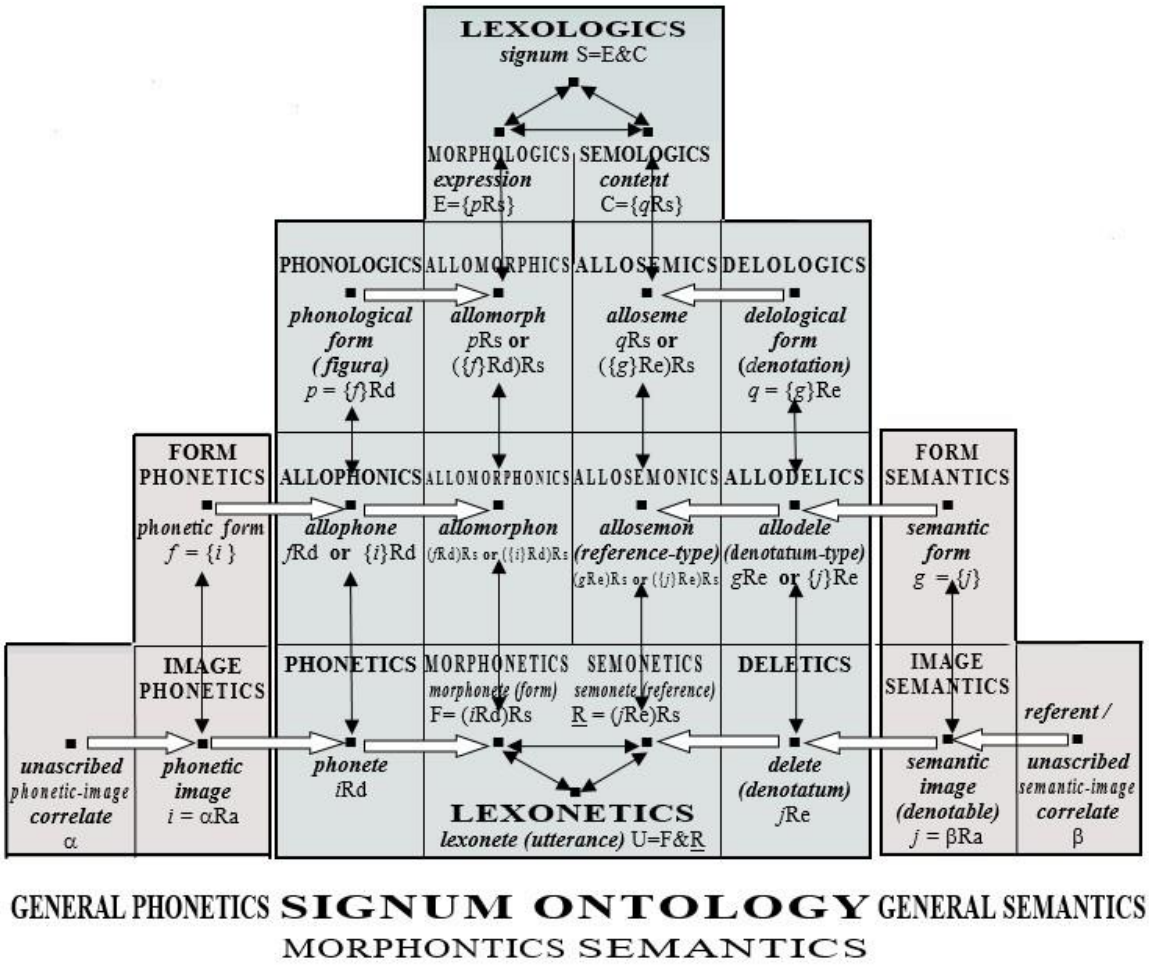

Figure 1: Extended axiomatic functionalism: signum ontology, plus general phonetics and general semantics

3. Proposed symbols for extended axiomatic functionalism

Figures 2–6 provide a list of proposed symbols for the major notions in the signum ontology and the ancillary areas of general phonetics and general semantics. These figures use for their illustrative examples, the English word ‘egg’. The material in figures 2–6 will be discussed in more detail in sections 4–4.5.

Figure 2: General-phonetic notions

Name	**Symbolisation of theoretical notion**	**Symbolisation for descriptive entity**	**Symbolisation of example of descriptive entity**
unascribed phonetic-image correlate	α	α *and* $^{\text{i, ii, iii, iv}}$, etc.	α^{i}
phonetic image	αR$\{^{\text{i…n}}\}$ *or i*	‹ ›*	‹εg›*
phonetic form	*f or* {*i*}	[ ]*	[εg]*

Figure 3: General-semantic notions

Name	**Symbolisation of theoretical notion**	**Symbolisation for descriptive entity**	**Symbolisation of example of descriptive entity**
unascribed semantic image-correlate / referent	β	β *and* $^{\text{i, ii, iii, iv}}$, etc.	β^{i}
semantic image	βR$\{^{\text{i…n}}\}$ *or j*	‹ ›*	‹*oval or round re-productive body …*›*
semantic form	*g or* {*j*}	[ ]*	[*oval or round re-productive body …*]*

Figure 4: Morphontic (non-meaning-related) only notions

Name	**Symbolisation of theoretical notion**	**Symbolisation for descriptive entity**	**Symbolisation of example of descriptive entity**
phonete	*i*Rd	‹ ›	‹εg›
morphonete	F *or* (*i*Rd)Rs	‹‹ ››	‹‹εg››
allophone	*f*Rd	[ ]	[εg]
allomorphon	(*f*Rd)Rs *or* {(*i*Rd)Rs}	[[ ]]	[[εg]]
phonological form / figura	*p or {f}Rd*	/ /$^{\text{SIG}}$ *or* / / (see Section 4.2)	/εg/
allomorph	*p*Rs *or* {({*f*}Rd)Rs}	// //	//εg//
expression	E *or*{*p*Rs}	⦃ ⦄$^{\text{E}}$	⦃egg⦄$^{\text{E}}$

Figure 5: Semantic (meaning-related) only notions

Name	**Symbolisation of theoretical notion**	**Symbolisation for descriptive entity**	**Symbolisation of example of descriptive entity**
delete/denotable	*j*Re	‹ ›	e.g. ‹*oval or round reproductive body …*›
semonete/reference	R *or* (*j*Re)Rs	‹‹ ››	‹‹*oval or round reproductive body …*››
allodele / denotatum-type	*g*Rr	[ ]	[*oval or round reproductive body …*]
allosemon / reference-type	(*g*Re)Rs *or* {(*j*Re)Rs}	[[ ]]	[[*oval or round reproductive body …*]]
delological form / denotation	*q or* {*g*}Re	/ /^{SIG} *or* / / (see Section 4.2)	/*oval or round reproductive body …*/
alloseme	*q*Rs *or* {({*f*}Re)Rs}	// //	//*oval or round reproductive body …*//
content	C *or* {*q*Rs}	⦃ ⦄^{C}	⦃*egg*⦄^{C}

Figure 6: Both morphontic (non-meaning-related) and semantic (meaning-related) notions

Name	**Symbolisation of theoretical notion**	**Symbolisation for descriptive entity**	**Symbolisation of example of descriptive entity**
lexonete/utterance	U *or* (*i*Rd)Rs & (*j*Re)Rs	‹‹ ››^{SIG}	‹‹egg››^{SIG}
signum	S *or* E&C *or* {*p*Rs}&{*q*Rs}	⦃ ⦄^{SIG}	⦃egg⦄^{SIG}

4. Exemplification

The notions of the signum ontology and the ancillary areas of general phonetics and general semantics are described in detail in Dickins (2009 and 2016). Readers unfamiliar with the theory are advised to read Dickins (2016) in particular before carrying on with this article. What follows is a brief discussion of the notions of general phonetics, general semantics and the system ontology. These are presented in the order in which they are given in figures 2–6. The notions of the signum ontology and the ancillary areas of general phonetics and general semantics can be illustrated on the basis of the English word (as a kind of signum) ‘egg’, which has a single allomorph of phonological form /ɛg/ and which has various senses (allosemes), such the one having the delological form / denotation /*oval or round reproductive body laid by the females of birds, reptiles , fishes , insects, and some other animals, consisting of a developing embryo , its food store, and sometimes jelly or albumen , all surrounded by an outer shell or membrane*/ (Collins English Dictionary); for the purpose of convenience of representation in this article, this can be shortened to /*oval or round reproductive body …*/.

For the representation of delological form / denotation as /*oval or round reproductive body …*/, see Section 4.1 below.

4.1 General-phonetic notions

Unascribed phonetic-image correlate

As a theoretical notion, unascribed phonetic-image correlate is symbolised as α (Figure 1, Figure 2). As a descriptive entity, an unascribed phonetic image-correlate can be symbolised as α^{i}, α^{ii}, α^{iii}, α^{iv}, etc, a complete set of unascribed phonetic image-correlates being symbolised as $\alpha^{\{i\ldots n\}}$ (Figure 2). An unascribed phonetic-image correlate is a "'propertiless' model for an individual real-world speech sound (uttered at a particular time and place). All that an unascribed phonetic-image correlate does is to identify this speech sound as existing" (see discussion of 'Peircean first' in relation to unascribed semantic-image correlate / referent, Section 4.2; also Dickins 2016: 17). Thus, if I utter a sound at 11.43 am on June 25, 2014 in room 4.05 in the Michael Sadler Building, University of Leeds, England, and note this as existing, this is an unascribed phonetic-image correlate.[4]

Phonetic image

As a theoretical notion, 'phonetic image' is symbolised as *i* and can be analysed as αRa (Figure 1, Figure 2), where a is an arbitrary set-forming criterion (for discussion of the difference between arbitrary and non-arbitrary set-forming criteria, see Dickins 2016: 15, 23, 26, 36, 37). As a descriptive entity, a phonetic image is symbolised using ‹ ›*, an example of a phonetic image being ‹ɛg›* (Figure 2). "Phonetic image provides a 'propertied' model for an individual speech-sound, occurring at a particular time and place, and thus gives us a basic model which we can use to describe the phonetic data" (Dickins 2016: 17). "It does not bear any relationship to a phonological entity / figura.[5] Thus, if I utter the sound 'ɛg' at 11.43 am on June 25, 2014 in room 4.05 in the Michael Sadler Building, University of Leeds, England, and I simply note this as a specific, individual speech sound, this – or rather the

[4] Reviewer 2 has questioned whether unascribed phonetic- and semantic-image correlates can be symbolized at all, on the grounds that if they are symbolized, they are assigned to some reality, be it only a particular graphic symbol, i.e. they gain some properties by this. I am not sure this is right, since the symbol stands *for* something, rather than representing or being a property of that thing. However, if it is right, I suggest this is better regarded as a paradox rather than a contradiction, i.e. something which we are better to live with as an 'irritant', rather than considering it to introduce an insoluble problem. Accordingly, we could think of the symbol as something which is there, but which ideally, if it were possible to symbolise without having a symbol, would not be there.

[5] In this article, and elsewhere (e.g. Dickins 2020), I have taken 'figura' and 'phonological form' to mean the same in relation to natural language. In fact, all the terminologically non-integrated terms in Figure 1 – 'utterance', 'form', 'reference', 'denotatum', 'denotable', 'referent', 'reference-type', 'denotatum-type', 'denotation', 'expression' and 'content', as well as 'figura' – can be used of both non-linguistic as well as linguistic semiotic systems. (There is a further issue, of whether 'figura' is to be taken to mean the same as 'cenological form' – i.e. phonological form, in relation to natural spoken language. This falls outside the scope of this article, but would need, ultimately, to be resolved; cf. Dickins 2009: 15, Def. 2b; 15, Def. 2b1; 15; Def. 2b1d; 35, Def. 23; and other definitions in which 'figura' and 'cenological form' occur.)

model for this – is the phonetic image ‹εg›* (cf. Dickins 2016: 21). See also Dickins (2009: 35, Def. 22) for a formal definition.

Phonetic form
As a theoretical entity, phonetic form is symbolised as *f* and can be analysed (defined) as {*i*} (Figure 1, Figure 2). As a descriptive entity, a phonetic form is symbolised using [ ]*, an example of a phonetic form being [εg]* (Figure 2). A phonetic form is a generalisation "from phonetic image to the entire set of phonetic images which are deemed identical apart from their time-space individuality (specificity) […]". Phonetic form provides the basic general model which allows us to describe speech sounds not simply as individual occurrences, but as more abstract generalised notions – e.g. the speech sound [εg]*, as a general notion, rather than simply a speech sound ‹εg›* which was uttered in a particular place at a particular time (cf. Dickins 2016: 17). See also Dickins (2009: 35, Def. 22a) for a formal definition.

4.2 General-semantic notions

Unascribed semantic-image correlate / referent
As a theoretical notion, an unascribed semantic-image correlate / referent is symbolised using β (Figure 1, Figure 4). As a descriptive entity, an unascribed semantic image-correlate / referent can be symbolised as β^{i}, β^{ii}, β^{iii}, β^{iv}, etc, the complete set being symbolised as $\beta^{\{i\ldots n\}}$ (Figure 4). An unascribed semantic-image correlate / referent is "a model for a 'propertiless meanable entity'; all that it involves is its mere existence. Unascribed semantic-image correlate / referent would appear to be very similar to a Peircean 'first' – and may, indeed, be exactly the same as a Peircean 'first' (cf. Gorlée, 2009)" (Dickins 2016: 15). An example of an unascribed semantic-image correlate / referent is a model for an egg (oval or round reproductive body …) without being ascribed to the category of egg (or any other category) (cf. Dickins 2016: 16).

Semantic image / denotable
As a theoretical notion, semantic image / denotable is symbolised as *j* and can be analysed as βRa (Figure 1, Figure 3), where a is an arbitrary set-forming criterion (for discussion of the difference between arbitrary and non-arbitrary set-forming criteria, see Dickins 2016: 15, 23, 26, 36, 37). As a descriptive entity, an example of a semantic image / denotable is ‹*oval or round reproductive body …*›* (Figure 4). Note that the same angle brackets, ‹ and ›, are used for semantic image / denotable as for phonetic image. The difference between them is marked by the fact that in the case of phonetic image, the element within these brackets is in plain font, while in the case of semantic image / denotable, it is in italics. The same distinction between the use of plain font and italics is made for all entities in general phonetics and general semantics (plain font for general-phonetic entities, italics for general-semantic entities), and also all entities within the morphontics and semantics of the signum ontology (plain font for morphontic entities, italics for semantic entities).[6]

[6] Reviewer 2 has noted that given the very frequent use of italics in general writing, the fact that these are the only way in which semantic are distinguished from morphontic entities might prove problematic for readers,

A semantic image / denotable is a model for a 'propertied' meanable entity, i.e. it is a meanable entity which is ascribed to (belongs to) a category (set) of meaningful entities. Thus, while a referent (in a particular case) is a model for an oval or round reproductive body … without being ascribed to the category of oval or round reproductive body … (or any other category), a semantic image (in a particular case) is a model for an oval or round reproductive body … which is ascribed to the category of 'oval or round reproductive body …'. Semantic image / denotable does not bear any relationship to a delological entity / denotation (cf. Dickins 2016: 16). See also Dickins (2009: 35, Def. 23b) for a formal definition.

Semantic form

As a theoretical notion, semantic form is symbolised as *g* and can be analysed as {*j*} (Figure 1, Figure 3). As a descriptive entity, a semantic form is symbolised using [ ]*, an example of a semantic form being [*oval or round reproductive body …*]* (Figure 3). A semantic form is a generalisation from semantic image to the entire set of semantic images which are deemed to belong to the same category (set) (cf. Dickins 2016: 17). Semantic form provides the basic general model which allows us to describe 'meanable' entities not simply as individuals, but as more abstract generalised notions – e.g. [*oval or round reproductive body …*]*, as a general notion, rather than simply a particular specific ‹*oval or round reproductive body …*›* (cf. Dickins 2016: 15). See also Dickins (2009: 36, Def. 23b1) for a formal definition.

4.3 Morphontic (non-meaning-related) only notions

Phonete

As a theoretical notion, phonete is symbolised as *i*Rd (Figure 1, Figure 4). As a descriptive entity, a phonete is symbolised using ‹ ›, an example of a phonete being ‹ɛg›, as an instantiation (individual realisation) of the phonological form / figura /ɛg/) (Figure 1, Figure 3). A phonete is a phonetic image *i* (or αRa) brought into a relationship R with a phonological distinctive function d, i.e. phonological entity / phono in the system ontology (see Section 6) (cf. Dickins 2016: 22). See also Dickins (2009: 45, Def. F3d) for a formal definition.

Morphonete/form

As a theoretical notion, morphonete/form is symbolised as F, and can be analysed as (*i*Rd)Rs (Figure 1, Figure 4). As a descriptive entity, a morphonete/form is symbolised using ‹‹ ›› ^{E}, an example of a morphonete/form being ‹‹ɛg››^{E}, as 'simultaneously' an instantiation (individual realisation) of the phonological form / figura /ɛg/ and the signum ⦃egg⦄^{SIG}) (Figure 4). A morphonete/form is a phonete *i*Rd brought into a relationship R with a grammatical distinctive function, i.e. grammatical entity / lexo in the system ontology (see Section 6) (cf. Dickins 2016: 33). See also Dickins (2009: 43, Def. F1b1a4) for a formal definition.

and perhaps even writers. I hope that this is not the case. However, it would need to be tested through the practical use of this symbolisation and consideration of reader reactions.

Allophone
As a theoretical notion, allophone can be analysed as *f*Rd or {*i*}Rd (Figure 1, Figure 4). As a descriptive entity, an allophone is symbolised using [ ], an example of an allophone being [ɛg], as the immediately generalised realisation of a phonological form / figura /ɛg/ (Figure 4). An allophone is a set of phonetes (a set whose members can each be analysed as *i*Rd) which are identical except for their time-space individuality (specificity) (Dickins 2016: 21). An allophone is also a phonetic form *f*, brought into a relationship with a phonological distinctive function d, i.e. with a phonological entity / phono in the system ontology (see Section 6). See also Dickins (2009: 35, Def. 23a1) for a formal definition.

Allomorphon
As a theoretical notion, allomorphon can be analysed as (*f*Rd)Rs or as ({*i*}Rd)Rs (Figure 1, Figure 4). As a descriptive entity, an allomorphon is symbolised using [[ ]], an example of an allomorphon being [[ɛg]], as 'simultaneously' the immediately generalised realisation of the phonological form / figura /ɛg/ and the signum ⦃egg⦄^{SIG} (Figure 4). An allomorphon is a set of morphonetes/forms which are deemed identical except for their time-space individuality (specificity). An allomorphon is also an allophone *f*Rd, brought into a relationship with a lexological distinctive function s, i.e. with a lexo / lexological entity in the system ontology (cf. Dickins 2016: 34). See also Dickins (2009: 37, Def. 24b1e) for a formal definition.

Phonological form / figura
As a theoretical notion, a phonological form / figura is symbolised as *p*, or can be analysed analytically as {*f*}Rd (Figure 1, Figure 4). As a descriptive entity, a phonological form / figura is symbolised using / /^{SIG} (or / /, see below), an example of a phonological form / figura being /ɛg/ ^{SIG} (Figure 3). A phonological form / figura is a set of allophones, each in a relationship with a phonological distinctive function d, i.e. with a phonological entity / phono in the system ontology (see Section 6) (cf. Dickins 2016: 32). See also Dickins (2009: 15, Def. 2b) for a formal definition.

There is a one-to-one correspondence between phonological form / figura in the signum ontology and phono / phonological entity in the system ontology, a phono / phonological entity being the d, which is the right-hand term (distinctive function) in the definition of a phonological form / figura. A particular phonological form / figura is accordingly an extensional entity to which the corresponding intensional entity is the phono / phonological entity which constitutes that phonological form / figura's right-hand term, d (cf. Dickins 2016: 39). Where it is not considered important to differentiate between a phonological form / figura, as an entity in the signum ontology, and its corresponding phono / phonological entity in the system ontology, the symbolisation / / without the following superscript ^{SIG} can simply be used.

Allomorph
As a theoretical notion, allomorph can be analysed as *p*Rs or ({*f*}Rd)Rs (Figure 1, Figure 4). As a descriptive entity, an allomorph is symbolised as // //, an example of an allomorph being //ɛg// (corresponding to the phonological form / figura /ɛg/ and as a realisation of the signum ⦃egg⦄^{SIG}) (Figure 4). An allomorph is a set of allomorphons, and a secondary generalisation, i.e. a set of sets (Section 5.3.3) of morphonetes/forms (cf. Dickins 2016: 16). See also Dickins (2009: 37, Def. 24b1a) for a formal definition.

Expression
As a theoretical notion, expression is symbolised as E, and can be analysed as {*p*Rs} (Figure 1, Figure 4). As a descriptive entity, an expression is symbolised as ⦃ ⦄^{E}, an example of an allomorph being ⦃egg⦄^{E} (Figure 3). An expression is a set of allomorphs (cf. Dickins 2016: 32). See also Dickins (2009: 36, Def. 24a) for a formal definition.

4.4 Semantic (meaning-related) only notions

Delete/denotatum
As a theoretical notion, delete/denotatum is symbolised as *j*Re (Figure 1, Figure 5). As a descriptive entity, a delete/denotatum is symbolised using ‹ ›, an example of a delete/denotatum being ‹*oval or round reproductive body …*›, as an instantiation (individual realisation) of the delological form / denotation /*oval or round reproductive body …*/ (Figure 5). A delete is a semantic image / denotable *j* (or βRa) brought into a relationship R with a delological distinctive function e, i.e. with a delological entity/ delo in the system ontology (see Section 6) (Dickins 2016: 24).

Semonete/reference
As a theoretical entity, semonete/reference is symbolised as R, and can be analysed as (*j*Re)Rs (Figure 1, Figure 5). As a descriptive entity, a semonete/reference is symbolised using ‹‹ ››^{C}, an example of a semonete/reference being ‹‹*oval or round reproductive body …*››^{C} (Figure 5), as 'simultaneously' an instantiation (individual realisation) of a content ⦃egg⦄^{C}) of the delological form / denotation /*oval or round reproductive body …*/ and the signum ⦃egg⦄^{SIG}. A semonete/reference is a delete/denotatum *j*Re brought into a relationship R with a grammatical distinctive function s, i.e. with a grammatical entity / lexo in the system ontology (see Section 6) (cf. Dickins 2016: 33) (cf. Dickins 2016: 33). See also Dickins (2009: 43, Def. F1b2a5) for a formal definition.

Allodele/denotatum-type
As a theoretical notion, allodele/denotatum-type can be analysed as *g*Re or as {*j*}Re (Figure 1, Figure 5). As a descriptive entity, an allodele/denotatum-type is symbolised using [ ], an example of an allodele/denotatum-type being [*oval or round reproductive body …*] (as the immediately generalised realisation of a delological form / denotation /*oval or round reproductive body …*/) (Figure 5). An allodele/denotatum-type is a set of deletes/denotata (a set whose members can each be analysed as *j*Re) which are identical except for their individuality (cf. Dickins 2016: 21, 25). An allodele/denotatum-type is also a semantic form *g*, brought into a relationship with a delological distinctive function e, i.e. with a delological entity / delo in the system ontology (see Section 6). See also Dickins (2009: 36, Def. 23c2) for a formal definition.

Allosemon/reference-type
As a theoretical notion, allosemon/reference-type can be analysed as (*g*Re)Rs or ({*j*}Re)Rs (Figure 1, Figure 5). As a descriptive entity, an allosemon/reference-type is symbolised using [[ ]], an example of an allosemon/reference being [[*oval or round reproductive body …*]],

as 'simultaneously' the immediately generalised realisation of the delological form / denotation /*oval or round reproductive body …*/ and the signum ⦃egg⦄^{SIG}) (Figure 5). An allosemon/reference-type is a set of semonetes/references which are deemed identical except for their individuality. An allosemon/reference-type is also an allodele/denotatum-type *g*Re, brought into a relationship with a lexological distinctive function s, i.e. with a lexo in the system ontology (cf. Dickins 2016: 38). See also Dickins (2009: 38, Def. 24c1e) for a formal definition.

Delological form / denotation

As a theoretical notion, delological form / denotation is symbolised as *q*, and can analysed as {*g*}Re (Figure 1, Figure 5). As a descriptive entity, a delological form / denotation is symbolised using / /^{SIG} (or / /, see below), an example of a delological form / denotation being /*oval or round reproductive body …*/ (Figure 3). A delological form /denotation *q* is a set of allodeles/denotatum-types, each in a relationship with a delological distinctive function e, i.e. with a delological entity / delo in the system ontology (see Section 6) (cf. Dickins 2016: 32). See also Dickins (2009: 16, Def. 2c1d) for a formal definition.

There is a one-to-one correspondence between delological form / denotation in the signum ontology and delo / delological entity in the system ontology, a delo / delological entity being the e, which is the right-hand term (distinctive function) in the definition of a delological form / denotation. A particular delological form / denotation is accordingly an extensional entity to which the corresponding intensional entity is the delo / delological entity which constitutes that delological form / denotation's right-hand term, e (cf. Dickins 2016: 39). Where is it not considered important to differentiate between a delological form / denotation, as an entity in the signum ontology, and its corresponding delo / delological entity in the system ontology, the symbolisation / / without the following superscript ^{SIG} can simply be used.

Alloseme

As a theoretical notion, alloseme can be analysed as *q*Rs or ({*g*}Re)Rs (Figure 1, Figure 5). As a descriptive entity, an alloseme is symbolised using // //, an example of an alloseme being //*oval or round reproductive body …*// (corresponding to the delological form / denotation /*oval or round reproductive body …*/ and as a realisation of the signum ⦃egg⦄^{SIG}) (Figure 5). An alloseme is a set of allosemons, and a secondary generalisation, i.e. a set of sets (Section 5.3.3) of semonetes/references (cf. Dickins 2016: 14). See also Dickins (2009: 37, Def. 24c1a) for a formal definition.

Content

As a theoretical notion, content is symbolised as C and can be analysed as {*q*Rs} (Figure 1, Figure 5). As a descriptive entity, a content is symbolised as ⦃ ⦄^{C}, an example of a content being ⦃*egg*⦄^{C} (Figure 5). A content is a set of allosemes (cf. Dickins 2016: 32). See also Dickins (2009: 37, Def. 24b) for a formal definition. Note that while elsewhere for semantic (content-side) entities relating to the signum ⦃egg⦄^{SIG}, I have used '*oval or round reproductive body …*', with the content itself, i.e. the semantic 'aspect' of ⦃egg⦄^{SIG} (the signum), I have used the form '*egg*' (i.e. in ⦃*egg*⦄^{C}). The reason for this is that, while all the realisational semantic notions (alloseme, allosemon/reference-type, semonete/reference, delological form

/ denotation, allodele/denotatum-type, and delete/denotatum) bear a clear relation to the semantic 'area' '*oval or round reproductive body …*', a content, as an 'aspect' of a signum does not really do so.

4.5 Both morphontic (non-meaning-related) and semantic (meaning-related) notions

Lexonete/utterance
As a theoretical entity, a lexonete/utterance is symbolised as U and can be analysed as F&R (Figure 1, Figure 6). As a descriptive entity a lexonete/utterance is symbolised using ‹‹ ››^{SIG}, an example of a lexonete/utterance being ‹‹ɛg››^{SIG} (Figure 4). A lexonete/utterance is a conjunction (biunity) of a morphonete and a semonete (cf. Dickins 2016: 28). See also Dickins (2009: 42, Def. F1b0b) for a formal definition.

Signum
As a theoretical notion, signum is symbolised as S, and can be analysed as E&C, i.e. the conjunction of an expression and a content (Figure 1, Figure 6). As a descriptive entity, a signum is symbolised as ⦃ ⦄^{SIG} (or ⦃ ⦄, see below), an example of signum being ⦃egg⦄^{SIG} (Figure 6) (cf. Dickins 2016: 8, etc.). See also Dickins (2009: 13, Def. 2a1) for a formal definition.

There is a one-to-one correspondence between signum in the signum ontology and lexo / lexological entity in the system ontology, a lexo / lexological entity being the s, which is the right-hand term (distinctive function) s in the definition of a signum. A particular signum is accordingly an extensional entity to which the corresponding intensional entity is the lexo / lexological entity which constitutes that signum's right-hand term, s (cf. Dickins 2016: 39). Where is it not considered important to differentiate between a signum, as an entity in the signum ontology, and its corresponding lexo / lexological entity in the system ontology, the symbolisation ⦃ ⦄ without the following superscript ^{SIG} can simply be used.

5. General principles

The following general principles have been used in drawing up the symbols in Section 3.

5.1 Retention of existing linguistic symbols

Where a well-established symbol already exists in linguistics (regardless of the precise definition in other approaches), e.g. [ ] for allophone, and / / for phonological entity, this has been retained within this symbology.

One exception is { }, which is fairly commonly used for 'morpheme' in other linguistic approaches. Since { } is used for 'set/class' in extended axiomatic functionalism, the use of { } for morpheme (or other types of signa) has been avoided here. I have, however, used ⦃ ⦄, which is reminiscent of the morpheme symbol { } for signa, expressions and contents, a morpheme being a type of these.

5.2 Use of identical symbols for morphontics (expression side) and semantics (content side), and plain vs. italic text

The number of symbols has been minimised by using the same symbols for both morphontic (expression-side) and semantic (content-side) notions. Thus both an allophone and an allodele / denotatum-type are symbolised by [ ], and both a phonological form / figura and delological form / denotation are symbolised by / /. In order to distinguish morphontic (expression-side) entities from semantic (content-side) entities, morphontic entities are represented in plain text, using IPA, or other appropriate, symbols, e.g. [ɛg] as an example of an allophone, or /ɛg/ as an example of a phonological form / figura. Semantic entities, by contrast are represented in italics, e.g. [*oval or round reproductive body ...*] as an example of an allodele / denotatum-type, and /*oval or round reproductive body ...*/ as an example of a delological form / denotation.

5.3 Degree of abstraction

A central criterion in the use of symbols is the degree of abstraction of the entity in question from linguistic reality, according to the four following levels.

5.3.1 Instantiation: i.e. direct model for speech events (level 1)

The entities at this level are:

unascribed phonetic-image correlate
phonetic image
phonete
morphonete/form
semonete/reference
delete/denotatum
semantic image/denotable
unascribed semantic-image correlate

The symbolisations of all entities at this level involve angle brackets ‹ and ›. The difference between the use of single angle brackets, ‹ and ›, and double angle brackets ‹‹ and ›› will be discussed in Section 5.4.

5.3.2 Immediate generalisation (sets) from speech events (level 2)

The entities at this level are:

phonetic form
allophone

allomorphon
allosemon/reference-type
allodele/denotatum-type
semantic form

The symbolisations of all entities at this level involve square brackets [ ]. The difference between the use of single square brackets, [and], and double square brackets, [[and]], will be discussed in Section 5.4.

5.3.3 Secondary generalisation (set of sets) from speech events (level 3)

The entities at this level are:

phonological form / figura
allomorph
alloseme
delological form / denotation

The symbolisations of all entities at this level involve forward-slanted brackets / /. The difference between the use of forward-slanted brackets, / and /, and double forward-slanted brackets, // and //, will be discussed in Section 5.4.

5.3.4 Tertiary generalisation (set of sets of sets) from speech events: signum level (level 4)

The entities at this level are:

signum
expression
content

The symbolisations of all entities at this level involve the 'fancy' brackets ⦃ and ⦄.

5.4 Degree of peripherality or centrality

The symbols used also reflect their degree of peripherality or centrality to linguistics. Apart from the most basic notions, unascribed phonetic-image correlate, and unascribed semantic-image correlate, symbols in the areas ancillary to the signum ontology, i.e. the symbols in general phonetics and general semantics, are all enclosed within a pair of single brackets, either ‹ › or [ ], with a superscript white star ☆ after them.

The symbols in the peripheral columns within the signum ontology, where the entities relate only to either distinctive function in phonology (d) or distinctive function in delology, (e) are enclosed in a pair of single brackets: either ‹ ›, [ ] or / /.

The symbols in the central columns of the signum ontology, where the entities relate to distinctive function in lexology (s) and also either to distinctive function in (d) or distinctive function in delology (e), are enclosed in a pair of double brackets: either ‹‹ ››, [[ ]] or // //.

6. System ontology

As noted in Section 1, extended axiomatic functionalism has two theoretical components, the signum ontology (plus ancillary areas of general phonetics and general semantics) and the system ontology. The system ontology can be represented as in Figure 7.

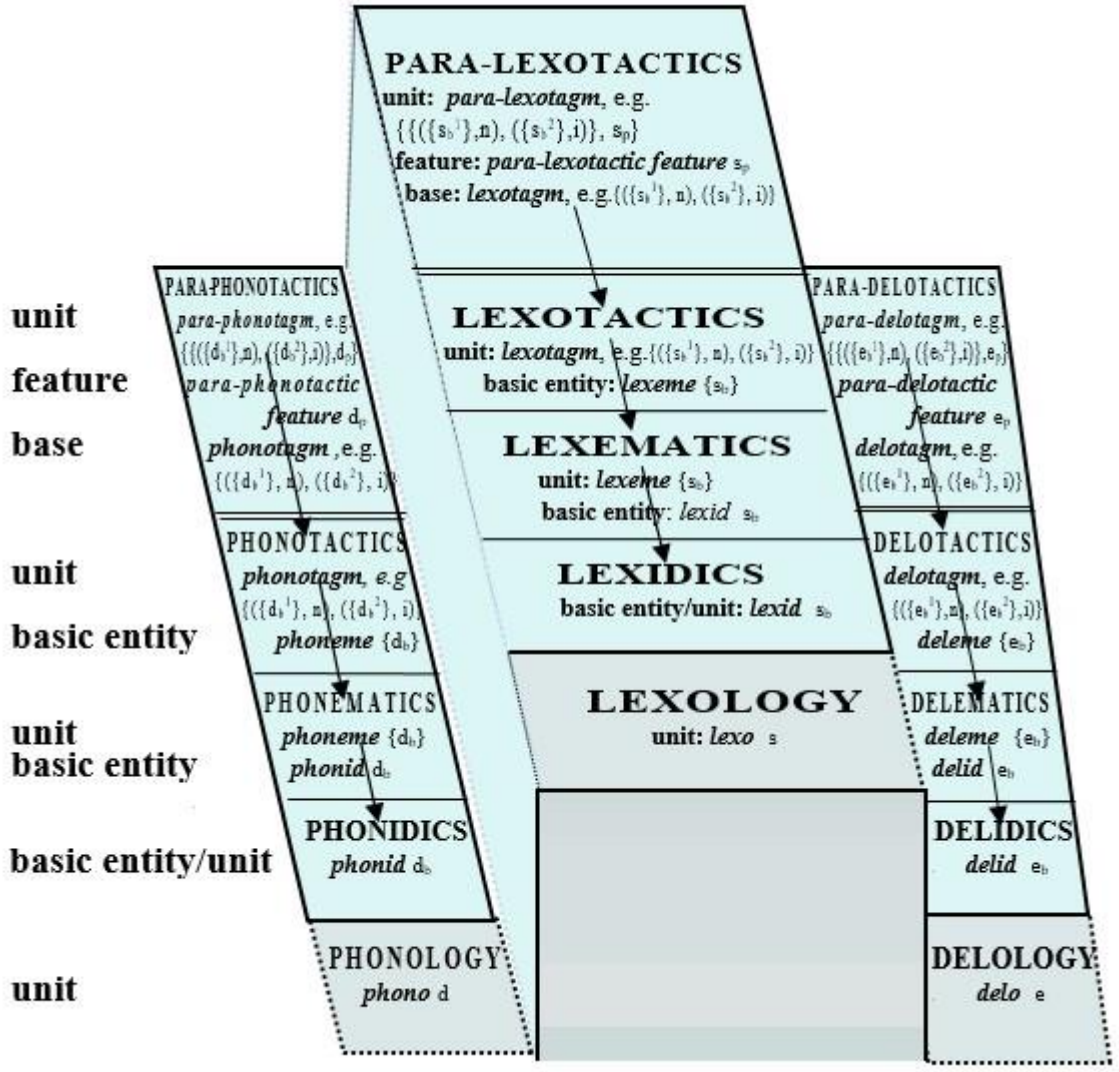

Figure 7: System ontology

Figure 8 (next page) shows the entire theory of extended axiomatic functionalism: the signum ontology (plus general phonetics and general semantics) and the system ontology.

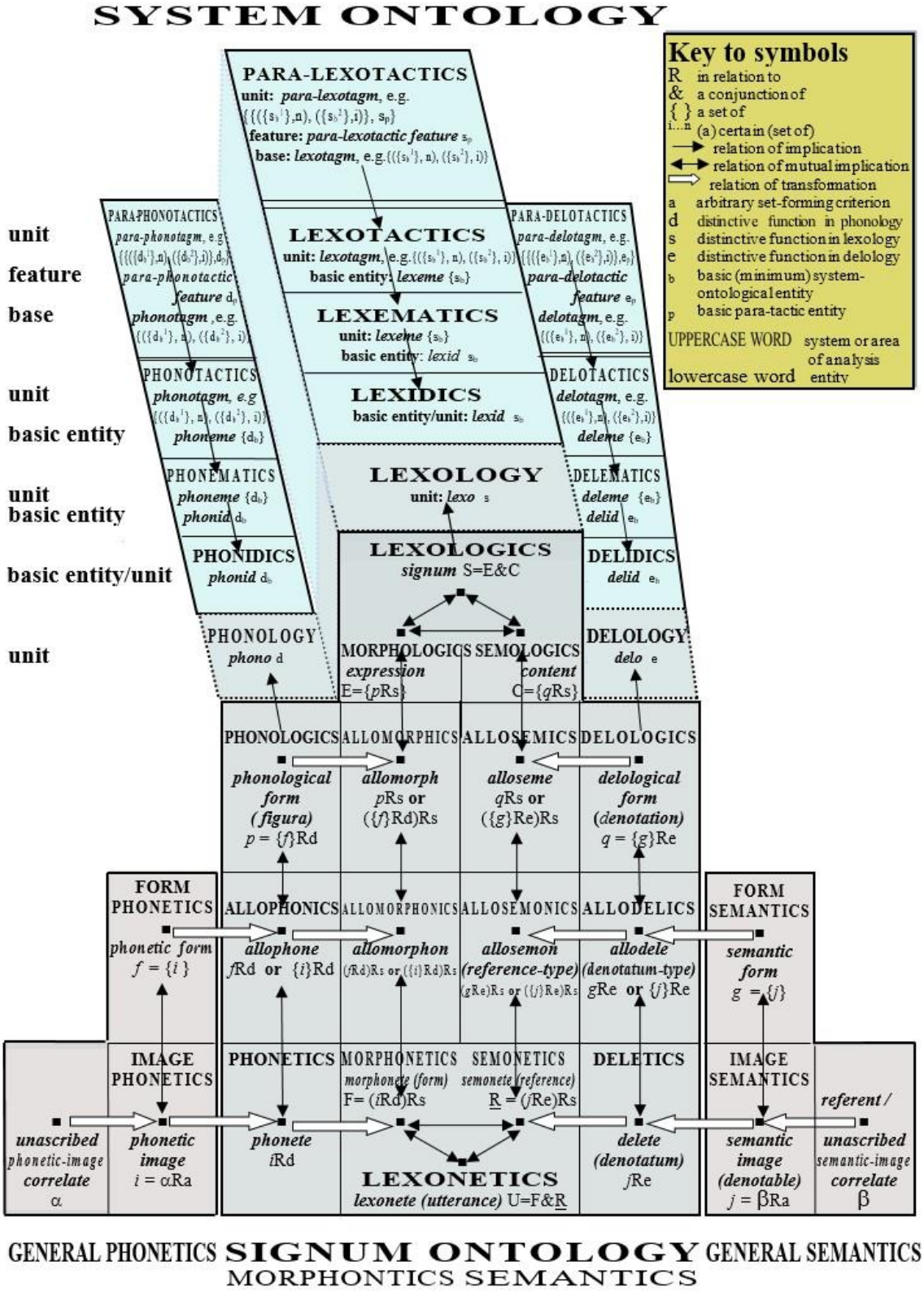

Figure 8: Extended axiomatic functionalism: signum ontology (plus general phonetics and general semantics) and system ontology

6.1 Proposed symbols for system ontology

The notions of the system ontology are described in detail in Dickins (2014 and 2016). Readers unfamiliar with the theory are advised to read Dickins (2014) in particular before carrying on with this article. As noted under the definitions in Figures 4 and 5 above, the entities in the system ontology – i.e. phonos / phonological entities, lexos / lexological entities, and delos / delological entities – are purely distinctive function (d, s and e). The system ontology comprises three sub-systems: phonology, lexology, and delology. For more detailed discussion of the system ontology, see Dickins 2009, 2014, and 2020).

Figure 9 re-presents the symbolisation for system-ontological entities already given in Figures 7 and 8, and provides a symbolisation for examples of phono, lexo and delo.

Figure 9: Notions in system-ontology (phonology, lexology and delology)

	Name	**Symbolisation of theoretical notion**	**Symbolisation of example of descriptive entity**
phonology	**phono (i.e. any phonological entity)**	d	/ɛg/^{SYS}
	phonid	d_b	**See Section 6.2**
	phoneme	$\{d_b\}$	
	phonotagm	e.g. $\{(\{d_b^1\}, n), (\{d_b^2\}, i)\}$	
	para-phonotactic feature	d_p	
	para-phonotagm	e.g. $\{\{(\{d_b^1\},n), (\{d_b^2\},i)\},d_p\}$	
lexology	**lexo (i.e. any lexological entity)**	s	⦃egg⦄^{SYS}
	lexid	s_b	**See Section 6.2**
	lexeme	$\{s_b\}$	
	lexotagm	e.g. $\{(\{s_b^1\}, n), (\{s_b^2\}, i)\}$	
	para-lexotactic feature	s_p	
	para-lexotagm	e.g. $\{\{(\{s_b^1\},n), (\{s_b^2\},i)\},s_p\}$	

delology	**delo (i.e. any delological entity)**	e	*/oval or round reproductive body … /*
	delid	e_b	**See Section 6.2**
	deleme	$\{e_b\}$	
	delotagm	e.g. $\{(\{e_b^1\}, n), (\{e_b^2\}, i)\}$	
	para-delotactic feature	e_p	
	para-delotagm	e.g. $\{\{(\{e_b^1\},n), (\{e_b^2\},i)\},e_p\}$	

6.2 Discussion of symbols for system ontology

In Figure 9, I have provided symbolisations for examples of a phono (i.e. any phonological entity – phonid, phoneme, phonotagm, or para-phonotagm), lexo (i.e. any lexological entity – lexid, lexeme, lexotagm, or para-lexotagm), and delo (i.e. any delological entity – delid, deleme, delotagm, or para-delotagm). These can be illustrated by /ɛg/^{SYS} as an example of a phono (more specifically a phonotagm), ⦃egg⦄^{SYS} as an example of a lexo (more specifically a lexeme), and /*oval or round reproductive body …* /^{SYS} as a delo (more specifically a delid, assuming this sense to be not further analysable delologically).

Just as slant brackets with plain text between them are used to symbolise phonological forms / figurae in the signum ontology, so slant brackets with plain text between them are used to symbolise phonos / phonological entities in the system ontology. Similarly, just as slant brackets with italic text between them are used to symbolise delological forms / denotations in the signum ontology, so slant brackets with italic text between them are used to symbolise delos / delological entities in the system ontology. Finally, just as fancy brackets with plain text between them are used to symbolise signa in the signa in the signum ontology, so fancy brackets with plain text between them are used to signify lexos / lexological entities in the system ontology.

Further, just as the symbolisation of phonological forms / figura, signa, and delological forms / denotations in the signum ontology involves a final superscript ^{SIG} (i.e. 'signum ontology) (Section 3), so a contrasting superscript ^{SYS} (i.e. 'system ontology') is used finally in the symbolisations for phonos / phonological entities, lexos / lexological entities and delos / delological entities in the system ontology. This allows for a clear and consistent differentiation between signum-ontological and system-ontological entities. Thus, /ɛg/^{SIG} is unambiguously a phonological form / figura (in the signum ontology), while the corresponding /ɛg/^{SYS} is unambiguously a phono / phonological entity (in fact a phonotagm) (in the system ontology).

Using relevant elements of the symbolisation in the 'symbolisation of theoretical notion' column, it would also be possible to construct more specific symbolisations for examples of specific descriptive entities (phonids, phonemes, phonotagms, and para-phonotagms; lexids,

lexemes, lexotagms, and para-lexotagms; and delids, delemes, delotagms, and para-delotagms). Thus, using the subscript $_{\text{b}}$, which symbolises the notion 'phonid' in the 'symbolisation of theoretical notion' column', we could symbolise the phonid (corresponding to what is traditionally termed a 'distinctive feature' in phonology) 'voiced' in Sudanese Arabic as /voiced/$_{\text{b}}^{\text{SYS}}$. Similarly, we could symbolise the phoneme 'p' in English as /{p}/$_{\text{b}}^{\text{SYS}}$ (using the { $_{\text{b}}$}, from the 'symbolisation of theoretical notion' column), and the phonotagm 'ɛg' in English as /{({ɛ}, n), ({g}, i)}/$^{\text{SYS}}$ (using the symbolisation {({ }, n), ({}, i)} from the 'symbolisation of theoretical notion' column).

Similarly, to take examples from lexology and delology, we could symbolise the lexeme 'egg' in English as ⦃{egg}⦄$_{\text{b}}^{\text{SYS}}$ (using the { $_{\text{b}}$} elements from the 'symbolisation of theoretical notion' column), and the deleme (assuming this in, in fact a deleme) 'oval or round reproductive body …' as /*oval or round reproductive body ...*/$_{\text{b}}^{\text{SYS}}$ (using the / /$_{\text{b}}$ elements from the 'symbolisation of theoretical notion' column).

This approach, however, would give rise to very cumbersome symbolisations, as can be seen, for example, of the symbolisation of the English phonotagm /{({ɛ}, n), ({g}, i)}/$^{\text{SYS}}$. Such symbolisations would also be unlikely, in practice, to add anything of value to existing, simpler alternative symbolisations. Thus, given a situation, within a piece of academic writing, in which we know that we are talking about the phonotactic analysis of 'ɛg', it is sufficient to use the simpler symbolisation {({ ɛ}, n), ({g}, i)}, merely using standard set-theoretical and relational notation, rather than the more complex /{({ɛ}, n), ({g}, i)}/$^{\text{SYS}}$, with an additional $^{\text{SYS}}$ at the end and slant brackets at the beginning and immediately before the $^{\text{SYS}}$, together with standard set-theoretical and relational notation.

7. Simplified symbolisation covering system ontology and signum ontology

A further simplification is achieved by the fact that in many cases (perhaps almost all cases), there will be no need to distinguish between an entity in the system ontology and its corresponding entity in the signum ontology. In such cases, the final final $^{\text{SYS}}$ (signifying a system-ontological entity) or $^{\text{SIG}}$ (signifying a signum-ontological entity) can simply be dropped. This would give, for example, /ɛg/, meaning either the phonological form / figura /ɛg/$^{\text{SIG}}$ in the signum ontology, or the corresponding phono / phonological entity /ɛg/$^{\text{SYS}}$ in the system ontology (or even, in some cases, both the signum-ontological phonological form / figura /ɛg/$^{\text{SIG}}$ and the corresponding system-ontological phono / phonological form /ɛg/$^{\text{SYS}}$, assuming these can be coherently talked about together in a particular context).

Correspondingly, in relation to lexology / the signum level, the non-necessity of using superscript $^{\text{SYS}}$ and $^{\text{SIG}}$, would mean that in appropriate contexts, ⦃egg⦄ could be used to symbolise either ⦃egg⦄$^{\text{SYS}}$ as a lexo / lexological entity (in fact. probably a lexeme) in the system onology, or ⦃egg⦄$^{\text{SIG}}$ as signum in the signum ontology (or even, if appropriate, both at the same time).

8. Comparison with Mulder's symbolisation for standard axiomatic functionalism

Figure 10 below compares the symbolisation for notions in extended axiomatic functionalism ('EAF' in Figure 10), as already discussed in this article with that given by Mulder for 'standard axiomatic functionalism' ('SAF' in Figure 10), i.e. the standard version of axiomatic functionalism developed by Mulder and Hervey (in Mulder 1989: 304). Figure 10 conforms to the overall shape of the signum ontology as given in Figure 1, but only includes those entities in the signum ontology of extended axiomatic functionalism for which there are equivalents in standard axiomatic functionalism and for which Mulder (1989: 304) provides symbols. In Figure 10, *x* is used within brackets, etc. to mean 'any entity'.

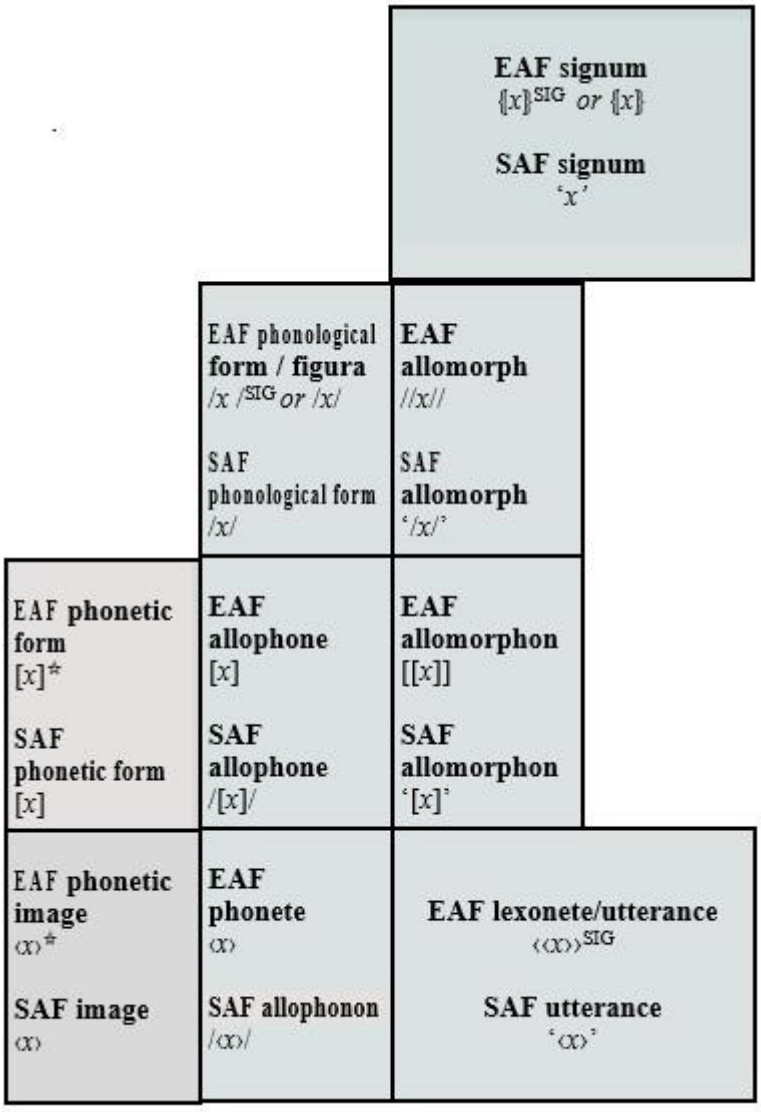

Figure 10: Comparison between proposed symbolisation for extended axiomatic functionalism and Mulder's symbolisation for standard axiomatic functionalism

Almost all notions have the same name in standard axiomatic as their correspondents in extended axiomatic functionalism. The only exceptions are standard axiomatic-functionalist

'image', which is termed 'phonetic image' in extended axiomatic functionalism, and standard axiomatic-functionalist 'phonete', which is termed 'allophonon' in standard axiomatic functionalism. 'Utterance' is used in both standard and extended axiomatic functionalism, but the latter also has the synonym 'lexonete'. 'Figura', which is a synonym of 'phonological form' in extended axiomatic functionalism, is a feature of the system ontology in standard axiomatic functionalism, and is therefore not included in Figure 10 (cf. Mulder and Hervey 2009: 3, Def. 2b, Def. 2b1).

The logic behind Mulder's proposed symbolisation for standard axiomatic functionalism is clear. Entities which directly model speech events (what I have termed in Section 5.3.1 'instantiations'; or 'level 1' entities) – i.e. images, allophonons and utterances – are symbolised using angle brackets ‹ ›. Entities which involve what I have termed 'immediate generalisations from speech events', also 'level 2' entities (Section 5.3.2) – phonetic forms, allophones and allomorphons – are symbolised using square brackets [ ]. Entities which involve what I have termed 'secondary' generalisations (i.e. sets of sets) from speech events', also 'level 3' entities (Section 5.3.3) – phonological forms and allomorphs – are symbolised using slant brackets. Entities which involve what I have termed 'tertiary' generalisations, also 'level 4' entities (i.e. sets of sets of sets), i.e. entities at the signum level (Section 5.3.4) are symbolised using single inverted commas: ‘ and ’.

Entities in standard axiomatic functionalism belonging to what I have termed 'general phonetics' (Figure 1) – i.e. images and phonetic forms – are symbolised using a single pair of brackets. Entities falling under the 'phonological form' column (what is termed in extended axiomatic functionalism 'phonotics'; Dickins 2009: 45, Def. F3f) – i.e. phonological form, allophone and allophonon – are symbolised using slant brackets / /. Thus, the symbol for a phonological form involves only slant brackets, it being both a secondary generalisation (set of sets) from speech events and falling under the 'phonological form' column. The symbol for an allophone involves slant brackets and square brackets (inside the slant brackets), since it is an immediate generalisation from speech events and falls under the 'phonological form' column. The symbol for an allophonon involves slant brackets and angle brackets (inside the slant brackets), since it is an instantiation of a speech event and falls under the 'phonological form' column.

Entities falling under the 'signum' column (what is termed in extended axiomatic functionalism 'lexotics'; Dickins 2009: 42. Def. F1b0d) – i.e. signum, allomorph, allomorphon and utterance – are symbolised using single inverted commas: ‘ and ’. Thus, the symbol for a signum involves only a single pair of single inverted commas, a signum being both a signum-level entity (level 4 entity) and falling under the 'signum' column. The symbol for an allomorph involves single inverted commas and slant brackets (inside the single inverted commas), since it is a level 3 entity and falls under the 'signum' column. The symbol for an allomorphon involves single inverted commas and square brackets (inside the single inverted commas), since it is an immediate generalisation from speech events (level 2 entity) and falls under the 'signum column. The symbol for an utterance involves single inverted commas and angle brackets (inside the single inverted commas), since it is an instantiation of a speech event (level 1 entity) and falls under the 'signum' column.

In terms of 'systemic consistency' of symbolisation, there seems little to choose between Mulder's symbolisation for standard axiomatic functionalism and the current proposed symbolisation for extended axiomatic functionalism – indeed Mulder's system might perhaps be

considered slightly more elegant and transparent. There are, however, a number of reasons why I think the symbolisation which I have proposed is better than that put forward by Mulder.

One of the principles which I identified for symbolisation (Section 5.1) was retention of existing linguistic symbols. As noted, this is achieved in the proposed symbolisation for extended axiomatic functionalism through the use of [x] for an allophone. In standard axiomatic functionalism, by contrast, an allophone is symbolised using the novel form /[*x*]/.

A further feature which I think is problematic in the symbolisation put forward by Mulder is the use of single inverted commas (‘ and ’) to symbolise a signum. Single inverted commas are used so frequently in general English writing (where they have a number of specific senses) that their use for symbolisation within a particular linguistic theory seems better avoided, and might, indeed, prove confusing in particular contexts. These problems are avoided in the use of ⦃*x*⦄, etc. in the symbolisation for extended axiomatic functionalism.

References

DICKINS, James. 2009. ‘Extended axiomatic functionalism: postulates’. *Linguistica* ONLINE. <http://www.phil.uni.cz/linguistica/art/dickins/dic-001.pdf>.

——. 2014. ‘On the non-necessity of levels in phonology, grammar and ‘abstract semantics’. *Linguistica* ONLINE. <http://www.phil.muni.cz/linguistica/art/dickins/dic-002.pdf>

——. 2016. ‘Construction of a linguistic theory from first principles and confrontation with crucial data’. *Linguistica* ONLINE. <http://www.phil.muni.cz/linguistica/art/dickins/dic-004.pdf>.

——. 2020. *Thematic structure and para-syntax: Arabic as a case study*. London and New York: Routledge.

GORLÉE, Dinda L. 2009. ‘A sketch of Peirce’s firstness and its significance to art’. *Sign Systems Studies* 37(1–2). Pp. 205–269.

MULDER, Jan A.F. 1998. *Foundations of axiomatic linguistics*. Berlin and New York: Mouton de Gruyter.

MULDER, Jan. A.F. and Sándor G. J. HERVEY. 2009. ‘Postulates for axiomatic functionalism’. *Linguistica* ONLINE. <http://www.phil.muni.cz/linguistica/art/mulderhervey/muh-001.pdf>.