# Sense Enumeration Vs Generative Lexicon Approach

***Dr. Sabri Elkateb***
***Dept. of English - Faculty of Education***
***Zawia University***

## Abstract:

One of the most striking problems facing theoretical and computational semantics is defining the representational link or interface between linguistic and non-linguistic knowledge. "Knowing a word is generally considered to be a matter of knowing the word's meaning, and meaning is one of those concepts of great importance for understanding the nature and limits of psychology" (Miller 1999). Lexical semantics is currently playing a crucial role in computational linguistics due to the fact that lexical entries in any representation must contain a considerable amount of information related to the word-sense. Massive research efforts have been so far directed towards an adequate approach or a model that introduces a knowledge representation framework which offers a rich and expressive vocabulary for lexical information. This paper is an attempt to

shed the light and draw a contrast on two opposed approaches to lexical semantic representation, namely: Sense Enumerative Lexicon (SEL) and Generative Lexicon (GL) approach. The main endeavor has been to investigate how adequate is each model to undertake extra intricate issues in lexical semantic representation.

# 1. Introduction:

Lexical representation is a commonly used term found in much recent work concerning the various areas of language processing, and has a longer history in the context of experimental and theoretical psycholinguistics. Most researchers in the field of language processing acknowledge that there is an intuitive sense of what 'lexical representation' refers to. According to Woollams (2015), lexical representation derives from our sense that there is some form of 'mental lexicon' or internal dictionary, in which the knowledge we have concerning the words we know is represented. Accordingly research has been apparently active to put forward a model that resembles somehow this human natural mental lexicon. An example of this is the Princeton WordNet which is mainly based on lexical Enumeration in its well formed representation. This paper discusses the Sense Enumerative Lexicon (SEL) approach as opposed to the Generative Lexicon (GL) Model in an attempt to figure out an acceptable model for lexical representation in both theory and practice.

# 2. Sense Enumeration Lexicon (SEL) Model:

According to (Pustejovsky, 1995:34) "a lexicon is a sense enumeration lexicon if and only if for "every word W in a language L, having multiple senses s1…,sn associated with that word, the lexical entries expressing these senses are sorted as {word sense (Ws1)………., word sense number (Wsn)}" This lexicon, termed by Pustejovsky (1995) as

"sense enumeration lexicon", lists all form–meaning associations in the language. A word form may be associated with one or multiple word meanings where it is said to be polysemous. A word meaning can be expressed or represented by more than one word form where these word forms are said to be synonyms.

A sense enumeration lexicon provides a clear framework for lexical representation in which the lexicon remains, as a component of a model of a speaker's linguistic competence, separate and independent from syntactic knowledge. This can be considered as a source of data from the point of view of computational linguistics. Sense enumeration lexicon adopts the standard lexical data structure of category type (CAT) and a basic specification of the genus (type) term (GENUS), as applied to the example below showing the structure of contrastive polysemy which can store any relevant information about each sense independently:

**bank$_1$**

$CAT = count\_noun$

$GENUS = financial\_institution$

**bank$_2$**

$CAT = count\_noun$

$GENUS = shore$

Further analysis of the senses of ***bank*** will be provided in subsection 2.2 below trough an actual extraction from the WordNet as a sense enumeration lexicon.

## 2.1 The WordNet as a Model for Sense Enumeration:

WordNet is a semantic lexicon for the English language developed by the linguist and psychology professor George Miller at Princeton University. WordNet is a sense enumeration lexicon which is considered one of the most important standard lexicons for English; WordNet aims at organizing, defining, and describing the relevant concepts of the English language. The main idea around it is the grouping of English words into sets of synonyms with same meanings called '*synsets'*. These sets or synsets are structured in ties and linked by meaning relations like hyponymy, antonymy, meronymy etc. Synonymy is a semantic relation between two words with different forms and similar meanings. According to Miller (1999), the traditional way to define synonymy is in terms of substitution: Two words are synonyms (relative to a context) if there is a statement (or class of statements) in which they can be interchanged without affecting truth value. Nouns participate in the relation of synonymy, antonymy, meronymy-holonymy, hypernymy-hyponymy relations, while verbs may be related by the synonymy, antonymy, troponymy, entailment, and hypernymy-hyponymy relations. Adjectives and adverbs are related by the synonymy and antonymy relations. That is to say, the lexicon is organised in terms of word meaning rather than word forms.

## 2.2. Solving the Problem of Polysemy in the WordNet:

Since many of the words are polysemous, morphological relations link synsets that have related meanings not individual words. For example, '*bank'* meaning (to do business with a bank) is linked to '*bank'* which means (depository financial institution), and bank meaning (to tip the aircraft laterally) should be linked to '*bank'* meaning (a flight manoeuvre), etc. "And in cases where the concepts of the noun and verb are different

e.g., womanize from woman no semantic link would need to be created" (Harabagiu et al 1999). The following example shows WordNet 2.1 sample result for derivationally related forms of the polysemous word '*bank*':

Figure 1. Extraction from WordNet 2.1 related forms of the polysemous word '*bank*'

**The noun bank has 10 senses (first 9 from tagged texts)**
1. **bank** -- depository financial institution, banking concern ………………
2. **bank** -- (sloping land (especially the slope beside a body of water ……………
3. **bank** -- (a supply or stock held in reserve for future use ………………
4. **bank** -- bank building -- (a building in which the business of banking transacted;…
5. **bank** -- (an arrangement of similar objects in a row or in tiers; …………
6. **bank** -- savings bank, coin bank, money box, (a container ………
7. **bank** -- (a long ridge or pile; "a huge bank of earth")………..
8. **bank** -- (the funds held by a gambling house or the dealer in some gambling games;
9. **bank** --  cant, camber (a slope in the turn of a road or track; ……………
10. **bank** -- (a flight maneuver; …………………

----------------------------------------------------------------

**The verb bank has 8 senses (first 2 from tagged texts)**
1. **bank** -- (tip laterally; "the pilot had to bank the aircraft")..........................
2. **bank** -- (enclose with a bank; "bank roads")………
3. **bank** -- (do business with a bank or keep an account at a bank;…………
4. **bank** -- (act as the banker in a game or in gambling)……..
5. **bank** -- (be in the banking business)………………..
6. **bank** -- deposit,  (put into a bank account; ………….
7. **bank** -- (cover with ashes so to control the rate of burning; "bank a fire") .................
8. **bank** --  trust, swear, rely, (have confidence or faith in; ……..

Further investigation of the senses in figure 1 above shows that
4 of 10 senses of bank as verb and as noun are semantically related:
**Sense** 1
**bank** --
depository financial institution, bank, banking concern, banking compa
ny -
(a financial institution that accepts deposits and channels the money int
o lending activities; "he cashed a check at the bank"; "that bank holds th

e mortgage on my home")
    RELATED TO->(verb) bank#3
     => bank --
(do business with a bank or keep an account at a bank; "Where do you b
ank in this town?")
    RELATED TO->(verb) bank#5
     => bank -- (be in the banking business)
    RELATED TO->(verb) bank#6
     => deposit, bank --
(put into a bank account; "She deposits her paycheck every month")

**Sense 2**
**bank** --
(sloping land (especially the slope beside a body of water); "they pulled
the canoe up on the bank")
    RELATED TO->(verb) bank#2
    => bank -- (enclose with a bank; "bank roads")

**Sense 4**
**bank**--
building (a building in which commercial banking is transacted; "the ba
nk is on the corner of.")
    RELATED TO->(verb) bank#3
    => bank --
(do business with a bank or keep an account at a bank; "Where do you
bank in this town?")
    RELATED TO->(verb) bank#5
    => bank -- (be in the banking business)

        RELATED TO->(verb) bank#6
     => deposit, bank --
 (put into a bank account; "She deposits her paycheck every month")

**Sense 10**
**bank** --
 (a flight manoeuvre; aircraft tips laterally about its longitudinal axis (es
pecially in turning);
"the plane went into a steep bank")
        RELATED TO->(verb) bank#1
        => bank -- (tip laterally; "the pilot had to bank the aircraft")

## 2.3. Merits of WordNet:

        WordNet as a lexical reference offers broad coverage of the general lexicon in English. WordNet has been employed as a resource for many applications in information retrieval. Knowledge of words lies not only in their meanings but also in the context in which they occur. Linking words to appropriate senses provides the desired conceptual information. Terms holding identical meanings are organized around the notion of a synset. Synsets are linked to each other via pre-defined lexical relations. Furthermore, WordNet's high level classes have put some limit to enumeration of word senses keeping limited the search space of any generalization process.

        Concepts are the organizational units in the WordNet and they are more than a single word as they include compounds, collocations, idiomatic phrases, and phrasal verbs. Jansen (2004) argues that compounds, collocations, idiomatic phrases, and phrasal verbs extend the idea of storing words in the lexicon to storing conceptual information that may not have a lexical representation using a single word.

## 3. The Generative Lexicon Model:

The central concern of Generative Lexicon theory is to minimize the need for enumerating word senses by providing operations for deriving and represnting most senses for a word from a basic one. This contrasts with sense-enumeration, in which several distinct senses are listed for a particular word.

The Generative Lexical model "has settled in the past years one of the most innovative prospective in lexical semantics". (Saint-Dizier, 1998) Aiming at laying the foundations of a theory of computational semantics, Pustejovsky (1991, 1995) outlines a conservative approach to decomposition, where lexical items are decomposed into structural forms or templates rather than sets of features. This model assumes that all lexical items are semantically active. The main idea of the generative lexicon model is that word senses are highly structured and the meaning of any word is not achieved by simply listing or enumerating its different senses. According to this approach the lexical item is viewed in the context rather than given an exhaustive description. The role played by all lexical items in the overall meaning of the sentence is highly emphasized.

## 3.1. Levels of Semantic Representation:

Four levels of semantic representation are put forward by Pustejovsky (1991, 1995) to characterize the system of generative lexicon. They are identified as Argument Structure, Event Structure, Qualia Structure and Lexical Inheritance Structure.

## a. Argument Structure:

A representation that determines a verb's meaning, defines the number and type of logical arguments and how they can be syntactically realized. A lexical item may have four distinct types of arguments:

- o **True argument**: parameters that are syntactically realized, e.g. 'John arrived late.'
- o **Default argument**: parameters that may not be syntactically expressed, but which participate in logical expression in the structural representation of the meaning of a lexical item, e.g. 'John built the house out of bricks.'
- o **Shadow argument:** parameters that can be found in the lexical item expressed by operations of subtyping or discourse, e.g., 'Mary buttered her toast with an expensive butter.'
- o **True Adjunct:** logical expressions modified by parameters that are part of the situational interpretation, e.g. 'Mary drove to New York on Tuesday.'

Pustejovsky (1995)

## b. Event Structure:

Event Structure is a representation that defines the type of event of a lexical item or a phrase. Event structure characterizes the event type of a lexical item and its internal structure. "A verb such as *build* involves a process and a resulting state." (Pustejovsky, 1995:71) Pustejovsky (1991) suggests that events are complex in the sense that they consist of subevents. Verbs and phrases containing verbs belong to either of three different types: they can be states, processes or transitions. Transitions are complex in that they consist of a process and a following state.

## c. Qualia Structure:

Qualia Structure essentially determines the meaning of a noun and deals with the different predications possible with a lexical item. It is the most innovative and important part of the generative lexicon model. Pustejovsky (1995) defines qualia structure as the structured representation which gives the relational force of a lexical item. Qualia are those aspects of word meaning that give a lexical item its role in an ontology. Qualia structure is claimed to be a system of relations that characterizes the semantics of nominals. It provides four essential aspects or roles of a word's meaning (or qualia) and they are what Pustejovsky termed *constitutive*, *formal*, *telic*, and *agentive* roles:

1. The formal role (i.e., hierarchical relations) refers to those features that distinguish an object within a larger domain, such as its orientation, magnitude, or shape which, for example, makes a *dictionary* identified as a kind of a *book*.

2. The constitutive role (i.e., meronymic relations) refers to the relation between an object and its constituents, such as its material, parts or components, and weight. In this case it indicates that a *dictionary* constitutes *information* (about words).

3. The telic role (i.e., functional) refers to the purpose and function of an item, such as the fact that a *dictionary* exists to be consulted while a *novel* exists to be read i.e. consulting a *dictionary* and reading a *novel*. Accordingly, the telic role represents the purpose behind performing an act and the function or aim which specifies certain activities.

4. The agentive role (i.e., construction/creation) encodes factors involved in the origin or bringing something about. For example, a *dictionary* comes about from compiling and a *novel* from writing.

Creator, Artefact, Natural kind, Causal chain. Pustejovsky (1991, **1995)**

To summarize the above four factors we could say, for example, for *novel*, the formal role is a kind of '*book*' that *constitutes* '*narrative*' and whose telic purpose is '*reading***'** of the thoughts which are brought about by the agent through '*writing*'. For more clarification, the above four aspects of qualia structure are illustrated in the diagram in Figure 8.1 showing a distinction between *novel* and *dictionary*:
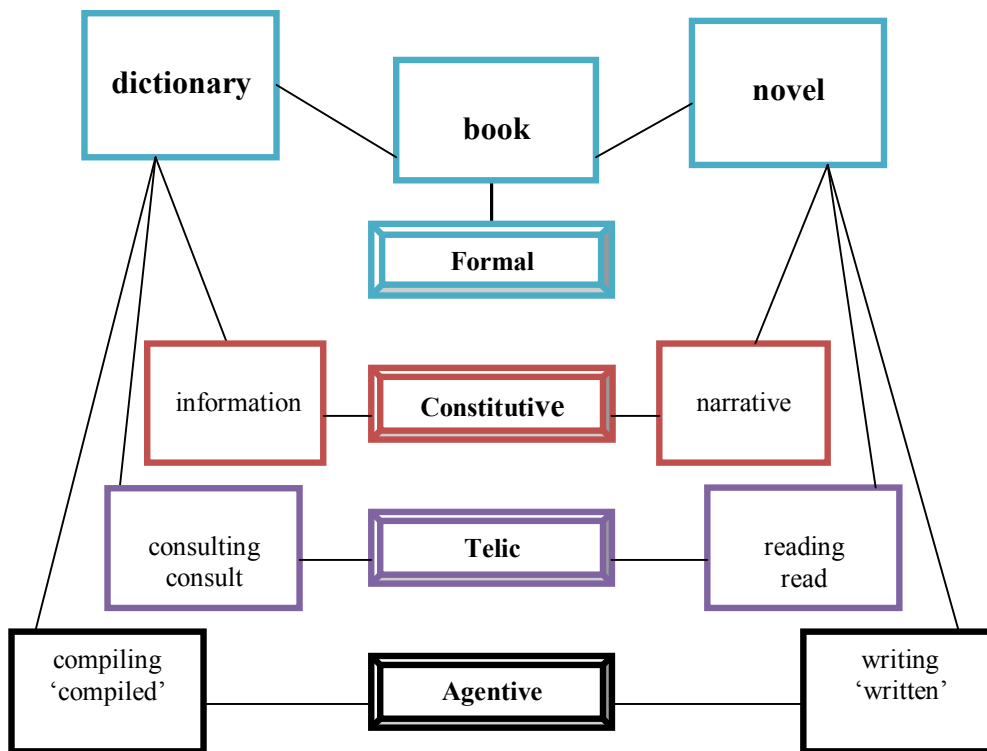
**Figure 2. The four aspects of qualia with the nouns 'novel' and 'dictionary'**

By integrating the four factors of qualia structure, Pustejovsky (1991, 1995) emphasizes that a lexical item contains much information that has been traditionally considered secondary or extra-linguistic in nature. "A close observation of linguistic data and of sense variations shows that the most important role is the telic role. It is in fact the role to be considered as a default role for the coercion object-event. The formal role is far less frequently considered." (Saint-Dizier, 1998:122)

## d. Lexical Inheritance Structure:

This level identifies how a lexical structure is related to other structures in the dictionary, or in the type lattice, however it is constructed. Lexical inheritance structure is said to achieve the following:

- o defines relations to other words in the lexicon.
- o provides a link to general world knowledge.

## 3.2. Implementation of Generative Lexicon:

Veale, T. (2003) describes how certain key elements of the qualia structure of a concept, pertaining to its agentive and telic properties can be automatically extracted from the sense glosses in WordNet. He claims that this approach to qualia extraction would balance coverage with quality. The extraction process deals primarily with a relatively narrow slice of the relational structure inherent in WordNet glosses. "Even this narrow slice yields a significant amount of qualia structure, since WordNet already encodes formal and constitutive relations in its taxonomic and meronymic links between synsets." (Veale 2003)

According to Veale (2003), the process operating entirely via a combination of derivational morphology rules and taxonomic sanity-checking, was able to automatically extract relationships from 40% of the noun glosses in WordNet 1.6. In addition, "96% of all noun glosses contain

at least one word with a denotation in the extraction set, which suggests that future extensions to the process may be able to obtain much higher coverage with relatively minor additions to the mechanism.

## 3.3. Arabic Morphology for Qualia Extraction:

In his attempt to extract the telic and agentive roles from the glosses of the WordNet Veale (2003) exploits the fact that the agentive and telic aspects of lexico-conceptual or morphosemantic structure are often expressed using nominalized verbs that implicitly encode relational structure. He states that a small number of highly productive morphology rules can thus be used to connect 'observe' to 'observer' and 'observation' (and vice versa), 'specialize', to 'specializer' and 'specialization', and so on. Thus, since WordNet 2.0 currently provides this feature that can link derivationally related lexical items with more precision through hand coding of nominalized matching pairs, more coverage of the process of qualia extraction became possible. It is worth attempting to demonstrate through some Arabic examples the capability of Arabic morphological structure to provide the required link between the above three items. This is also important for checking the possibility of extracting qualia from Arabic compared to some English counterparts. In this case, Arabic roots and patterns can be used to derive all three components. The number of patterns used can provide almost the full coverage of derivationally related words that form the required sets. Verbs and their corresponding derivatives are shown in Table 1 below.

**Table 1. derivationally related words as basis for qualia extraction**

| English Words | Word Pattern | root | Arabic Words |
|---|---|---|---|
| **(1)** | **(fa ?ala) — (fa :?il) — (fa ?l)** | | n    n    v |
| drive – driver — driving | sa :qa — sa :iq — sawq | s w q | ساق - سائق - سوق |
| move – mover — moving | naqala — na:qil — naql | n q l | نقل - ناقل - نقل |
| kill – killer — killing | qatala — qa:til — qatl | q t l | قتل - قاتل - قتل |

| **(2)** | **(fa ?ila) – (fa :?il) – (fi ?l)** | | |
|---|---|---|---|
| know – knower — knowing | ?alima - ?a:lim - ?ilm | ? l m | علم ــــ عالم - علم |
| keep – keeper — keeping | Hafizha — HafizH — Hifzh | H f zh | حفظ- حافظ - حفظ |

| **(3)** | **(fa ?ala) – (fa :?il) – (fi/a :lah)** | | |
|---|---|---|---|
| Plant – planter — planting | zara?a — zari? — zar? | z r ? | زرع ـــزارع ـــزراعة |
| write – writer- writing | kataba — ka:tib – kita:bah | k t b | كتب- كاتب- كتابة |
| make – maker — making | Sana ?a —Sa :ni ? — Sina :?ah | S n ? | صنع - صانع - صناعة |

| **(4)** | **(fa ?ala) – (fa :?il) – (fu ?u :l)** | | |
|---|---|---|---|
| arrive – arriveer– arrive ing/al | waSala – wa:Sil – wSu:l | w S l | وصل- واصل- وصول |
| fall – faller – falling | saqTa – sa :qiT – suqu :T | s q T | سقط- ساقط- سقوط |
| leave – leaver- leaving | Kharaja – kharij – khuru:j | kh r j | خرج- خارج- خروج |

| **(5)** | **(fa ??ala)– (mufa :??il)– (taf ?u :l)** | | |
|---|---|---|---|
| inspect – inspector — inspection | fattash — mufattish — tafti:sh | f t sh | فتش- مفتش- تفتيش |
| instruct – instructor – instuction | ?allama — mu ?allim — ta ?li :m | ? l m | علم- معلم- تعليم |
| falsify – falsifier – falsification | zawwara – muzawwer- tazwi:r | z w r | زور- مزور- تزوير |

| **(6)** | **(fa :?ala) (mufa :?il) (mufa :?alah)** | | |
|---|---|---|---|
| depart – departer – departure | Ga;dara — muGa:dir – muGa:darah | G d r | غادر- مغادر- مغادرة |
| support – supporter – support | sa :nada — musa :nid – musa :nadah | s n d | ساند- مساند- مساندة |
| gamble – gambler – gambling | qa :mara- muqa :mer- muqa :marah | q m r | قامر - مقامر- مقامرة |
| observe – observer – observation | ra :qaba – mura :qib – mura :qbah | r q b | راقب ــ مراقب ـ مراقبة |

The above Arabic sets show similar lexicalization to those in English and that each set can be derived using the roots and a small number of patterns that can be combined in a normal query without resort to morphological rules. Nominalizations do not seem to cause a problem for both derivation and qualia extraction for English and there is no possibility they might do for Arabic. The problem seems to underlie some nouns like

'*botanist*' and '*philologist*' in both languages. According to Veale (2003) WordNet concepts like these two are defined with glosses that explicitly employ the term "*specializing*", thus evoking the concept '*specializer*' (a hyponym of '*expert*') He argues that because '*specializer*' is compatible with the concepts '*botanist*' and '*philologist*' by virtue of being a hyponym of '*person*', this in turn suggests that '*botanist*' and '*philologist*' should be seen as hyponyms of '*specializer*', making *specializer_of* an appropriate telic relation for each. For Arabic the same can be applied for the above and similar cases using same relation '*mutakhaSSiS_fi*' (specializer_of) : '*takhaSSUS*' (specialization).

The Generative Lexicon formalism provides an innovative way of representing the semantics of objects and actions. Although it is developed for computational linguistics it contains some important features that could be useful in the dictionary. This model provides criteria for the description of a lexical item, most important of which is the *qualia structure* which describes how an entity is structured and what type of operations it can be involved in. We demonstrated how derivational morphology can put the basis for the extraction of the qualia (telic and agentive roles) from the Arabic glosses following the steps of Veale (2003) Once that more storage of Arabic words, roots, patterns and glosses is done the process of qualia extraction would be more fruitful.

## 3.4. Limitation of the Generative Lexicon Model:

Part of the representation is redundant with syntax. Since event structure composition is productive and does not actually need to be memorised, it is not clear whether it really belongs in a designated module sep- arate from syntactic generative devices proper, i.e. it looks like these principles need to apply to 'constructions' (Ramchand, 2006).

Contrary to Pustejovsky (1995) who claims that all structures according to his approach are productive, a structure as in (*John began a book*.) where it is not possible to predict that *John* began writing the book, or began reading it. Therefore, and according to Ramchand (2006), the effects of Qualia Structure are not distinguishable from real world knowledge. Thus, a structure like '*John began a book*' is not generative, or even predictable.

## 4. Conclusion:

Despite the, somehow, adequate results obtained by a model like the Princeton WordNet in several semantic and sense relations, there are still some limitations to the approach of SEL. This promising model has to be able to develop a complete range of usages and relations for a lexical item. The idea of a generative lexical model, discussed above, is contrasted to a more usual sense enumerative lexicon where each word has a literal meaning, and lexical ambiguity is treated by multiple listing of words.

In favour of the SEL approach, WordNet and similar projects like the EuroWorNet for Europian lamguages provide reliable models for such an approach of SEL. WordNet's high level classes have put some limit to enumeration of word senses keeping limited the search space of any generalization process. Concepts are the organizational units in the WordNet and they are more than a single word as they include compounds, collocations, idiomatic phrases, and phrasal verbs.

For the Generative lexicon approach we have attempted to follow the steps of Veale (2003) where we have demonstrated how derivational morphology can put the basis for the extraction of the qualia (telic and agentive roles) from Arabic according to the word and its root and pattern. Once that more storage of Arabic words, roots, patterns and glosses is done the process of qualia extraction would be more fruitful.

## 5. References:

1) *Black, W. J., and Elkateb, S. (2004)* **A Prototype EnglishArabic Dictionary Based on WordNet**, *Proceedings of 2nd Global WordNet Conference, GWC2004, Czech Republic, 67-74.*

2) *Elkateb, S and Black, W. J. (2001)* **Towards the Design of English-Arabic Terminological Knowledge Base**, *Proceedings of ACL 2000, Toulouse, France:113-118*

3) *Elkateb, S. & Black (2004) A* **Bilingual Dictionary with Enriched Lexical Information**, *Proceedings of NEMLAR Cairo, Egypt 2004 Arabic Language Tools and Resources:79-84*

4) *Elkateb, S., Black. W., Rodriguez, H, Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., (2006). Building a WordNet for Arabic, in Proceedings of the Fifth International Conference on Language Resources and Evaluation, Genoa, Italy.*

5) *Elkateb, S. (2014)* **The Need for a Reliable Dictionary**, *Zawia University Bulletin – ISSUE No.16- Vol. (4)*

6) *Elkateb, S. (2014)* **Words, Concepts and Meaning Representation**, *Zawia University Bulletin – ISSUE No.16- Vol. (4)*

7) *Harabagiu, S. Miller, G. and Moldovan, D. (1999)* **WordNet2 - a morphologically and semantically enhanced resource**. *In Proceedings of SIGLEX-99, pp. 1-8, University of Mariland*

8) *Jansen, P. (2004)* **Lexicography in an Interlingual Ontology**: *An Introduction to EuroWordNet, Canadian Undergraduate Journal of Cognitive Science, 2004 vol ii:1-5*

9) *Miller, G. A. (1990)* **Nouns in WordNet**: *A Lexical Inheritance System, Princeton University.*

10) *Miller, G. A. (1999) ON KNOWING A WORD, Annual Review of Psychology,* **50***, 1-19.*

11) *Nirenburg, S. & Raskin, V. (2001)* **Ten Choices for Lexical Semantics***, Computational Intelligence International Journal 17 (1): 157-177*

12) *Pustejovsky, J. (1991) The Generative Lexicon, Computational Linguistics 17 (4) 409-442.*

13) *Pustejovsky, J. (1995) The Generative Lexicon, Massachusetts Institute of Technology.*

14) *Ramchand, G. (2006)* **Against a Generative Lexicon,** *University of Geneva*

15) *Saint-Dizier, P. (1998)* **Sense Variation and Lexical Semantics Generative Operations***, In D.M.W. Powers (ed.) NeMLaP3/CoNLL98: New Methods in Language Processing and Computational Natural Language Learning, ACL:121-130.*

16) *Veale, T. (2003)* **Qualia Extraction from WordNet***, In the Proceedings of the 3ʳᵈ International Workshop on Creative Systems,*

*held as part of IJCAI'03, the 2003 International Joint Conference on Artificial Intelligence*

17) *Yule, G. (2010)* ***The Study of Language****, Cambridge University Press*

18) *Woollam, A.M.(2015)* ***Lexical is as lexical does: computational approaches to lexical representation****, Lang Cogn Neurosci. 2015 Apr 21; 30(4): 395–408.*