*Perspectives in Education 2014: 32(3)*
*ISSN 0258-2236*

*http://www.perspectives-in-education.com*

# Towards a reconceptualisation of "word" for high frequency word generation in word knowledge studies

**Jabulani Sibanda**

**Jean Baxen**

*The present paper derives from a PhD study investigating the nexus between Grade 4 textbook vocabulary demands and Grade 3 isiXhosa-speaking learners' knowledge of that vocabulary to enable them to read to learn in Grade 4. The paper challenges the efficacy of the four current definitions of "word" for generating high frequency words (HFW) for such a study. The paper submits that the token and type conceptualisations are unrealistic in their disregard of the learning burden principle, and that the lemma and word family notions are too accommodative and untenable in their over-extension of the learning burden principle. It critiques the arbitrary generation of word levels from a language corpus which is not cognisant of the natural order in which second language learners at different levels and from different first language backgrounds acquire English vocabulary. Based on research findings, the paper proposes, for the larger study, a unit-of-word quantification broader than the token but less accommodative than the lemma. The paper advocates further research into children's psychological processing of English word forms to constitute a taxonomy of word forms which merit treatment as single words at different levels of learners' competence.*

**Keywords:** vocabulary, high frequency words, type, token, lemma, word family

## Introduction

This paper interrogates the suitability of current "word" conceptualisations for a larger study determining the vocabulary knowledge of Grade 3 isiXhosa speaking learners in relation to the vocabulary needs of Grade 4 content area textbooks. A

**Jabulani Sibanda**
**Education Department,**
**Rhodes University**
**Email:jabusbnd@gmail.com**
**Telephone: 084 528 2087**

**Jean Baxen**
**School of Education, Wits University**
**E-mail: jean.baxen@wits.ac.za**
**Telephone:** 011 717 3014

determination of Grade 4 textbook vocabulary needs necessitates the generation of high frequency words (HFW) in the textbooks, without whose knowledge learners would not be able to read to learn. Even for English first language users, HFW have been perceived as a notorious obstacle (Morimoto & Loewen, 2007). Because these words are high coverage words in texts (Gardner, 2007) representing a high percentage of the total number of words met receptively and used productively (Hunt & Beglar, 2005), learners need to master them to register success in reading. Crossing that HFW threshold, positions children for better textual comprehension (Nation, 2001) and eases mastery of the less common words from context (Llach & Gallego, 2009). The study the present paper straddles Grade 3 and 4, a transitional stage from learning to read (mastery of the mechanical aspects of reading) to reading to learn (using the reading skill to access information in texts). In the study, the HFW in Grade 4 textbooks form the baseline against which third graders' vocabulary knowledge is measured. The generation of the HFW list is itself dependent on the conceptualisation of the construct "word" since different conceptualisations yield different word lists. This paper interrogates these perspectives, and proffers not only an alternative conceptualisation of "word" for the study, but also a way forward that takes account of variables and considerations which current perspectives seem to be overlooking.

## Conceptualisation of the construct "word"

The question "What is a word?" has plagued the field of vocabulary testing for years and has defied singularity or uniformity of definition. Discrepancies in vocabulary size estimates stem from a lack of consensus on what constitutes "word" as word frequency counts are contingent upon the word conceptualisation used as the unit of analysis. Knowing all the words in The singer sang a song different from the songs the other singers had sung can be regarded as knowing 14, 12, 9 or 7 words if "word" is defined as token, type, lemma or word family, namely the four current word conceptualisations, respectively. D'Anna, Zechmeister and Hall's (1991: 111) question, "[W]hen we say that a child learns 3,000 or 5,000 words per year, what exactly are we talking about?" is as valid now as it was then. Whether the four conceptualisations emerged consecutively or concurrently is not clear from literature and no researchers or authors are credited with particular perspectives.

This paper questions the efficacy of each of the four word conceptualisations and discusses their implications for the generation of English HFW in Grade 4 textbooks to be tested on second language Grade 3 users of English. An exposition of the limitations of the four perspectives on "word", which is the next subject of discussion, explains the need for an alternative perspective which this paper submits, as well as a way forward.

## The perspective of "word" as token

In the perspective of "word" as token, "… each occurrence of a form is counted separately" (Luitel, 2011: 59). The token yields the total quantity of textual input in raw terms (Mármol, 2011). According to Nation (2001), tokens are the unit we refer to when we talk about a summary, a telegram, or a research paper being so many words long. Tokens identify words "… simply by the space between the strings of letters in written language" (Luitel, 2011: 59). Carter (1998: 4) in Catalán and Francisco (2008: 151) defines a token as "… any sequence of letters (and a limited number of other characteristics such as hyphen and apostrophe) bounded on either side by a space or punctuation mark". Any expression devoid of spaces within it and separated by spaces from other expressions is consistent with the view of word as token.

There are, however, permutations and nuances to consider before adopting this concept of word for HFW computation purposes. Such a definition fails to account for some compound constructions. Closed-form compounds are coalesced together as in "childlike", hyphenated compounds are separated by a hyphen as in "four-year-old", and open compounds are separated as in "post office". Should hyphens then be considered as spaces or not? The inconsistency in the division of compounds such as "injustice" and "in-laws" is problematic when one considers hyphens as spaces. One could also ask: "Does the fact that an ice cream is one item make it one word or does the presence of two forms make it two words?" Mármol (2011) sees such as single words by virtue of their representing a single concept causing children to understand them as one concept.

A conceptualisation of word as token makes computation of word frequencies impossible, because every stand-alone form is regarded as diverse from the others. For vocabulary knowledge measurement, it is not plausible to consider the same word with the same orthographic make-up, meaning and pronunciation as a separate word every time it recurs in text. Therefore, an alternative way of conceptualising word-for-word frequency counts is needed, and word as type, which represents a moderate and less radical shift from the token conception than lemma and word family, has been put forward, and is considered next.

## The perspective of "word" as type

In word as type "… only the word form that is 'different' from all the remaining ones is counted" (Read, 2000 in Luitel, 2011: 59). A recurring word form is counted only once and so, although there are ten tokens in the statement, Your mother was talking to my mother in your garden, there are only eight types, because the words "your" and "mother" appear twice but are counted once each. Type as a unit of quantification assumes that a word's identity rests on its orthographic make-up where all words that are identically spelt are considered one word, an assumption which is not accommodative of homonyms such as the overused but apt example

of "bank". Words which function as both nouns and verbs depending on their use such as "pin" in the statement Get the pin and pin the papers on the board are not considered in word as type. Knowledge of "pin" as a noun does not guarantee that of "pin" as a verb. Being spelt the same does not translate to a word being the same word wherever it is encountered. The perspective of word as type also disregards the fact that the meanings of some words can be extrapolated from knowledge of related others. Knowing the word "boys" logically presupposes knowledge of the word "boy" and the two would well be considered as one word even for Grade 3 isiXhosa-speaking children. This is what Nation (2001) refers to as the learning burden principle which posits that one's knowledge of some words reduces or eliminates the burden that is required to learn or know other words. Both the token and the type formulations ignore this significant principle which is at the core of word acquisition and learning. The remaining conceptualisations, lemma and word family, whose limitations are discussed later, apply this principle, the lemma in not as broad a way as does the word family, which merits its discussion next.

## The perspective of "word" as lemma

The lemma is preferred for lexical quantification on account of overcoming the limitation of having to consider each word form as a unique form unrelated to the other forms, as does the type and token conceptualisations. Gardner (2007: 243,244) quotes Francis and Kucera's (1982: 1) definition of a lemma as "… a set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling". The verbs "write", "writes", "writing", "written" and "wrote" emanate from the base form "write" and so are all considered as one word. The inflections "-s", "-ing", "-en" do not create new words but merely modify the form in which the word "write" occurs to indicate grammatical functioning. The lemma is, therefore, premised on the assumption that, once the base form is known, the knowledge of the inflected forms is eased and expedited; the learning burden principle. Browne, Cihi and Culligan (2007: 2) rationalise the confinement of members of a lemma to a single word class by positing that the "… statistical item difficulty factors for 'accept', 'accepts' and 'accepting' are very close, whereas the statistical difficulties for 'acceptable', 'acceptance' and 'unacceptable', are all quite different. One hypothesis is that the brain treats these six items as four different Base Words." While the above example fits the argument, one might argue that knowledge of the base form "write" could make the form "writer" easier to one learner than the form "wrote" or "written" which belongs to the same word class as "write".

The accommodation of irregular words such as "mice", "beaten", "brought" and "best" within a lemma makes the assumption that belonging to the same word class as the base reduces the learning burden of a word highly suspect. As Gardner (2007: 244) observes:

> *the case of the irregulars poses serious quandaries relating to the psychological validity of such family relationships – namely, that the opaque spelling and*

*phonological connections between the lemma headword and the family members will surely cause more and different learning problems than their more transparent counterparts.*

This defeats the learning burden principle which the lemma should uphold and hence Nation's (2001: 8) question: "Should the irregular forms be counted as a part of the same lemma as their base word or should they be put into separate lemmas?" The orthographic constitution of the word "best", for instance, is not in any way indicative of stemming from the base form "good". The lemma should be a grouping of words whose understanding is made almost obvious whenever the base form is known and not words brought together by virtue of being inflected from the same base form. The criteria of lemma membership are not tight enough to ensure only those words whose meanings are easily recoverable from the meaning of the base gain entrance into the lemma.

Broadening the scope of a lemma to include the contracted forms as Nation (2001) does, introduces at least two challenges. First, knowledge of the contracted form requires knowledge of not only the base form, but also that of "not" since the contracted form is both a fusion and reduction of two words; a base form and "not", for example "can + not = can't". Second, what one would consider opaque contractions cannot easily be inferred from the base form + "not". Transparent contractions would be forms such as "have + not = haven't", "do + not = don't", and the opaque forms would be "will + not = won't", am + not = ain't", "shall + not = shan't". The opaque contractions have a higher learning burden than the transparent forms and assuming that L2 beginners can associate such irregular forms with their headwords is fundamentally unrealistic.

In Milton's (2009:10) definition a lemma "... includes a headword and its most frequent inflections and this process must not involve changing the part of speech from that of the headword". In formulaic terms this translates to:

*Lemma = headword + most frequent inflections + their contracted forms (belonging to same class).*

The phrase "most frequent" leaves the determination of most frequent to the researcher's discretion in the absence of frequency lists of inflected forms. The word "frequency" also needs qualification, whether it is the frequency with which the inflected form is used in a text or the frequency which stems from the number of English words that an inflection inflects. The former would be relative to text, because frequent forms in one text might be less frequent in another which renders it not tight enough to allow easy and objective application. The latter does not guarantee that inflections with a lower spread in their use are more difficult than those that have an impact on many word forms in the language.

The lemma also assumes that inflections are easier than other forms of affixation (prefixation and suffixation) which is controversial. Some suffixes such as "-able" and "-less" and prefixes like "-un" have meaning in and of themselves which can be used

to recover the meaning of a suffixed and prefixed form like "suitable", "careless" and "unfair", yet inflections are devoid of such independent meaning. Such systematic use of affixes can significantly reduce the learning burden of the words derived from a known base form. That the inflections "-s" and "-es" can be used for both verb and plural forms, could be a confounding factor on its own.

There is also a challenge of determining what really counts as a headword. The base form might not be the most common form or the one that the learners are likely to acquire first, which justifies Sinclair's (1991) supposed complication of which to consider as the headword, namely the base form or the most common form. Nation (2001) notes that, although the comparative and superlative forms have always been considered English inflections in the computerised lemmatised list of the Brown Corpus (Francis & Kucera, 1982), they are excluded in these definitions of lemma. This, however, speaks to the difficulty of constituting a lemma with precision. A proposal by Stubbs (2002) to have all members of a lemma share the same meaning fails to distinguish a lemma from a lexeme as the latter also denotes a group of words sharing the same meaning and the same word class. Acknowledging the difficulty of constituting a lemma and "… generalizations about whole lemma …", Knowles and Mohd Don's (2004: 71) advise for researchers, namely "… to consider individual words" as basis for word count and analyses, is almost a call to revert to word as type.

Insights from brain research support the learning burden principle but not the constitution of lemmas. Browne et al. (2007: 2) assert that "… the brain stores and processes lemmas having similar difficulty factors as forms of the same word, and … stores and processes lemmas having different difficulty factors as different words." Impressive formulaic definitions of lemma divorced from the psychological realities of learners' word learning do not advance the cause of vocabulary measurement. Browne et al.'s (2007) observation that some lemmas are registered by the brain as separate words rather than one casts doubt on the cogency of lemmas as units of frequency counts. That the brain does not always store and process lemmas as we constitute them, points to the need for a more tenable unit of word quantification. The inability of the three perspectives to define "word" with precision necessitates a consideration of the last unit word as word family.

## The perspective of "word" as word family

The word family construct comprises a headword, its inflected forms and closely derived forms (Nation, 2001) and can be represented thus:

*Word Family=Base form +Basic Inflected forms+ Transparent derivatives.*

Whereas inflection does not produce separate words, derivation creates separate but morphologically related words which usually involve some change in form. How to determine, with objective certainty, the basic inflected forms and the closely derived forms or transparent derivatives, is questionable. Word family unit assumes the application of morphological rules through affixation (prefixation and suffixation)

to word learning which ensures "… little or no extra learning when one or more of the members is already known to the learner" (Chung, 2009: 162). By including a "… wider range of inflections and derivations …" (Milton, 2009: 11) than the lemma and having members traverse word class boundaries, the word family unit extends the application of the learning burden principle further than the lemma does. From the base form "long", the forms "longer", "longest", "longish", "length", "lengthen" and "lengthy" could be considered as one word despite the apparent diversity in the learning burdens of these forms. Even derived forms differ in their complexity (Browne et al., 2007).

The efficacy of the word family unit is challenged by Mármol (2011: 12) who says:

> *… we cast doubt on the idea that a child acquiring bed has also acquired bedroom. There is the possibility that an adult could guess the meaning of the latter, but a young language learner in his first stages of acquisition may not be able to make those inferences.*

Word family unit assumes an intricate knowledge of morphological inflections of the English language in order. Evidently, isiXhosa-speaking Grade 3s would not possess the native-like intricate knowledge of morphological relations between words needed to make intelligent guesses about the meaning of some words on the basis of knowledge of their base form. In a study which required non-native postgraduate and undergraduate participants to identify the derivational forms of stimulus stem words, Schmitt and Zimmerman (2002) found that participants could only rarely provide all the different derivations of the stimulus words. This indicated partial knowledge of derivational forms on the part of the participants. The complexity of morphological knowledge that learners should possess to be able to process words as word families is illustrative in their need to know that "mean" does not derive from "me" despite the orthographic string for "me" occurring in "mean" (Bauer & Nation, 1993). Implicit knowledge of the role of affixes in word formation and meaning, which is requisite for the use of word family unit, should not be assumed especially for L2 users of English.

From studying 1 000 000 token Lancaster-Oslo-Bergen corpus, Bauer and Nation (1993) came up with a seven-level scheme for defining word families "… at various defensible levels for analysis and comparative analysis purposes – at least in terms of learners' abilities to associate morphologically related words" (Gardner, 2007: 247). The scheme is based on an analysis of inflections and affixations of English words based on their productivity, frequency, regularity and predictability. Their taxonomy, which represents a scaling of word families from the most elementary and transparent members to those of less obvious possibilities (Nation, 2001), is an acknowledgement that word families are not a one-size-fits-all phenomenon. Table 1 adapted from Bauer and Nation (1993: 254) shows the seven levels of inflections and affixations.

***Table 1: Bauer and Nation's (1993) word family levels***

| *Level* | *Affixation and inflection* |
|---|---|
| 1 | No affixes |
| 2 | -s, -ing, -ed, -er, -est (all inflections) |
| 3 | -able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un- (most frequent and regular derivational affixes) |
| 4 | -al, -ation, -ess, -ful, -ism, -ist, -ity, -ise, -ment, -ous, in- (frequent, orthographically regular affixes) |
| 5 | -age, -al, -ally, -an, -ance, -ant, -ary, -atory, -dom, -eer, -en, -ence, -ent, -ery, -ese,<br>-esque, -ette, -hood, -i, -ian, -ite, -let, -ling, -ly, -most, -ory, -ship, -ward, -ways, -wise, ante-, anti-, arch-, bi-, circum-, counter-, en-, ex-, fore-, hyper-, inter-, mid-, mis-, neo-, post-, pro-, semi-, sub-, un- (regular but infrequent affixes) |
| 6 | -able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-, re- (frequent but irregular affixes) |
| 7 | ab-, ad-, com-, de-, dis-, ex-, sub- (classical roots and affixes) |

For a learner operating at level 5, for instance, words in levels 1 to 5 emanating from the same base would be considered as one word, but those in levels 6 and 7 would be regarded as different words from the base form. Although rigorous criteria (productivity, frequency, regularity and predictability) were applied to a large corpus (a million words), the levels generated could be questioned. First, is the repetition of many affixed forms at the different levels. The suffix "-able" belongs to both level 3 and 6 which presents uncertainty about membership level of forms such as "suitable" on the scale. Second, is the failure to acknowledge the different learning dilemmas that derivational affixes might pose to developing readers. Third, is the assumption that both exposure to and acquisition of morphologically related words follow a linear pattern where base forms are acquired prior to their inflected and derived forms. The assumption is refuted by Biemiller and Slonim (2001) who note that young children might actually acquire many derived forms before they acquire their root-form counterparts. It is suspicious for a form such as "disadvantage" (level 7) according to the taxonomy to have a lower learning burden than "advantageous" (level 4) when the opposite seems more likely. It is problematic to base word family levels solely on a corpus without empirical evidence of the ease with which learners acquire the different affixed forms. Seen in this light, Bauer and Nation's (1993) work

requires further large-scale research to corroborate the match between the levels of the corpus analysis and the psychological realities of learners' word learning and acquisition.

This critique shows that the learning burden principle is not judiciously considered by all four word constructs. The token and type disregard their application on the one hand by assuming that each form is acquired and known independently from the other forms, which necessitates their being considered as a separate word. The lemma and word family on the other hand over-extend their application by considering as one word, several words whose relationship to the base might not even transparent or apparent. The result, then, are four word constructs which scarcely respond to the challenge the large study confronts, namely a unit of word quantification consonant with Grade 3 isiXhosa-speaking learners on the verge of a transition to learning through the medium of English in Grade 4. This paper argues for a conceptualisation of the construct "word" which is accommodative of diverse learner competencies and linguistic backgrounds among other variables.

## "Taxonomy of word for word counts": A potential way forward

Among the disadvantages that L2 English learners face are: no tacit knowledge of English which L1 learners possess, limited initial linguistic resource base, pressure to learn to read English simultaneously with the acquisition of the language's oral language which the L1 learners already possess, and fundamental differences between their L1 and English particularly at the lexical level. None of the current word notions conceptualise "word" in a plausible way for the generation of HFW for L2 English Grade 3s.

This paper proposes two ways of dealing with the challenge:

1. An alternative conceptualisation of word as a unit of counting for the larger study. Words sharing the same base and the following inflections: the present progressive (-ing) as in "eating", plural (-s) as in "books", possessive (-'s) as in "boy's", past regular (-ed) as in "talked", third person singular (-s) as in "walks", and the long plural (-es) as in "mangoes" are considered as one word and all other forms as separate words. Krashen (1987) reports on extensive morpheme studies by Dulay and Burt (1974), Fathman (1975) and Makino which showed that these inflections are the ones that learners naturally acquire first, and in the order in which they are listed independently of instruction, learners' age, L1 background, or conditions of exposure. This would recognise the learning burden principle unlike the token and type perspectives, but not over-extend the principle by being overly accommodative like the lemma and type perspectives.

2. While the alternative above is accommodative of the context of the larger study on which this paper is based, there is need for a taxonomy where the above conceptualisation of word would be but one level out of several others. The taxonomy that this paper proposes should have word levels corresponding to the natural order in which particular word forms are acquired once the most common form is known and not on the basis of the structure of a language corpus. Such a hierarchy would be at the intersection between Bauer and Nation's (1993) word family levels which give a hierarchy of the relative ease of acquisition of particular words, and morpheme studies which generated a list of the natural order of the acquisition of English grammatical structures. The one is based on a study of the language and the other has its basis on a study of the learners' "natural order" of acquiring words. These two would determine the extent of the match between the language corpus ideals and the psychological realities of the learners' acquisition of words.

The paper also argues for large-scale testing and documentation of the order of acquisition of English affixed forms (suffixed and prefixed for both inflections an derivations) on learners from particular language backgrounds, including that of L1 English users. Such testing needs to be done at varying levels of learner competence and with all the affixed forms of words. Brown's (1973) longitudinal study, reported in Kwon (2005: 4), produced the order of L1 acquisition of English morphemes which was unfortunately confined to morphemes and based exclusively on English native users. A tenable word construct should ensure that only word forms deriving from the base which are regarded as the same word with the base, pose negligible or no learning burden in the event that the base form is known. If research determines that in the natural order of vocabulary acquisition the form "writing" comes way earlier than "wrote" then, at an early level of the taxonomy, "write" and "writing" could be regarded as one word with "wrote" being considered one word with "write" at a later level. There would then be just a taxonomy with several levels at which word is defined and from which researchers can conceptualise word depending on their purposes and contexts.

The token would form level 1 of the taxonomy, as every form is considered a separate word. Studies not generating HFW lists would use this level. The second level would be what we presently have as the type. Again, this level would not be any use in word frequency counts, because it ignores the learning burden principle. Level 3 could be the alternative notion of word proposed in 1 above, straddling token and lemma. Several levels would egress in what could possibly be called "taxonomy of word conceptualisation". Base forms which are acquired earliest and the words emanating from the base which children demonstrate proclivity to acquire first no matter their word classes would be considered. Specific rules would then be worked for word membership at each level and exceptions noted. Lexicographers would make dictionary entries of the word conceptualisation level for each form. After defining "boys", an addition can be made plural form of "boy" [l.2] for level 2. If

in doubt one would reference a word's level from the base form. In their studies, researchers would specify levels at which they based their definition of word as a unit of measurement.

The proposed taxonomy would allow the replication of studies, as one can tell with certainty which words in the study were used as one word. Teachers would select vocabulary for instructional purposes in a way that facilitates and expedites the natural order of the acquisition of particular words. The taxonomy would also ensure that the affixed, inflected and derived forms learners demonstrate proclivity to acquire with ease at their different levels as L1 and L2 speakers of English are used to assess their word knowledge. It would also further our knowledge of the psychological realities of children's acquisition of English words, which brain research can take forward. A challenge, though, would be coming up with rules governing forms to consider as one word at each level as well as identifying all the possible exceptions to the rules. The absence of particular rules might make the taxonomy a little messy but might not be a bad thing considering that some of the limitations of the lemma and word family can be traced to a desire to have objective rules governing membership into those "families". Further debate on these proposals would potentially lead to a refinement of these ideas and, ultimately, the creation of more useful word construct for word frequency counts.

## References

Bauer L & Nation ISP 1993. Word families. *International Journal of Lexicography*, 6: 253-279.

Biemiller A & Slonim N 2001. Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology,* 93(3): 498-520.

Browne C Cihi G & Culligan, B 2007. Measuring vocabulary size via online technology. Lexxica. Retrieved from http://www.wordengine. jp/research/main.

Catalán RJ & Francisco RM 2008.Vocabulary input in EFL textbooks. *RESLA*, 21: 147-165.

Chung TM 2009. The newspaper word list: A specialised vocabulary for reading newspapers. *JALT Journal*, 31(2): 159-182.

D'Anna CA, Zechmeister EB & Hall JW 1991. Toward a meaningful definition of vocabulary size. *Journal of Literacy Research*, 23(xxiii): 109-122.

Francis W & Kučera H 1982. *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Gardner D 2007. Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28(2): 241-265.

Hunt A & Beglar D 2005. A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, 17: 23-59.

Knowles G & Mohd Don Z 2004. The notion of a "lemma": Headwords, roots and lexical sets. *International Journal of Corpus Linguistics*, 9(1): 69-81.

Krashen SD 1982. *Principles and practice in second language acquisition*. Oxford: Pergamon.

Kwon EY 2005. The "natural order" of morpheme acquisition: A Historical survey and discussion of three putative determinants. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 5(1): 1-21.

Llach MPA & Gallego MT 2009. Examining the Relationship between Receptive Vocabulary Size and Written Skills of Primary School Learners. J*ournal of theSpanish Association of Anglo-American Studies*, 31(1): 129–147.

Luitel B 2011. Vocabulary in the New B.Ed. General English under Tribhuvan University. *Nepal English Language Teachers' Association Journal of NELTA,* 16(1-2): 59-69.

Mármol A 2011. Vocabulary input in classroom materials: Two EFL course books used in Spanish schools by Gema. *RESLA*, 24: 9-28.

Milton J 2009. *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Morimoto S & Loewen S 2007. A comparison of the effects of image-schema-based instruction and translation-based instruction on the acquisition of L2 polysemous words*. Language Teaching Research*, 11(3):347-372.

Nation ISP 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University. Schmitt N & Zimmerman CB 2002. Derivative word forms: What do learners know? *TESOL Quarterly,* 36(2): 145-171.

Sinclair J McH *1991. Corpus, concordance, collocation. In Stubbs M. (1996). Text and Corpus Analysis. Oxford, UK: Blackwell.*

Stubbs M 2002. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishing.