

# Register Variation and the Multi-word Item

Alexandra Uzoaku Esimaje

Faculty of Arts and Education, Benson Idahosa University, Benin City, Edo state, Nigeria

Email: alexandra.esimaje@live.com; sandra.esimaje@gmail.com

**Abstract**—Past research in language has posited numerous reasons for language variation. The findings of such research seem to have focused more on how language varies in different contexts, where contexts cover factors of geography, ethnicity, social class, age, sex and culture on the one hand, and situational variables such as mode, domain and tenor on the other. But beyond these, a new paradigm has emerged with the advent of computer-assisted language research, which has primed investigations into similarities and dissimilarities between register types based on various linguistic phenomena. One of the central issues in such research is lexical behaviour, which is the main interest of this research, in particular, the use of multi-word items which are postulated to characterize individual registers. This study focuses on this relatively new area of linguistic enquiry, to account for how a text type can be distinguished based on the word sequences which compose it. The question at issue is – are there genre-specific multi-word items? To answer this question, the paper investigates the use of multi-word items in five different texts: fictional writing, academic writing, religious speech, political speech, and memorial speech. The results show some indication that the multi-word item can serve as an index of text variability, but precisely which multi-word items are expected to occur in a register, or not occur, remains inconclusive. We therefore recommend further and continual research into the use of this item until a complete picture emerges.

**Index Terms**—registers, language variation, multiword items, lexical behaviour

## I. INTRODUCTION

In general usage in linguistics, a text is a passage of any length which forms a unified whole. It can be in oral or written form. However, not all written or oral matter qualifies as text. A text is a unit of language in use. It is not a grammatical unit, like a clause or sentence; and it is not defined by its size. As Halliday and Hasan (1976, 1989) explain, a text is sometimes envisaged to be some kind of super sentence, a grammatical unit that is larger than a sentence but related to a sentence in the same way that a sentence is related to a clause, a clause to a group and so on: by constituency. A text is not something that is like a sentence, only bigger; it is something that differs intrinsically from a sentence. 'A text is best regarded as a semantic unit: a unit not of form but of meaning. Thus it is related to a clause or sentence not by size but by realization, the coding of one symbolic system in another. A text does not consist of sentences; it is realized by or encoded in, sentences' (Halliday & Hasan, 1976, p. 3-4). A text is: an instance of language in use, either spoken or written; a piece of language behaviour which has occurred naturally without the intervention of the linguist, e.g. a conversation, lecture or sermon. Every text has a property or properties which qualify it as text. This is called texture; these are linguistic features which unify it as text, creating the cohesion necessary for its understanding. In linguistic study, some terminological confusion attends 'text' and 'discourse'. It is needful to indicate how text relates to discourse, at least in this study.

Baker (2006) discusses the various ways that discourse has been used in social and scientific research: as language above the sentence level, as language in use, as types of language use or topics, and as practices which systematically form the objects of which they speak. Of these inter-related yet different conceptualisations, two of the senses which this study adopts are the first and second ones; discourse is language above the sentence level and language in use. According to Stubbs (1996), it is naturally-occurring connected speech or writing; the organisation of language above the sentence or above the clause, which follows that discourse, is language use in social contexts, the interaction or dialogue between speakers. Therefore, discourse analysis is: analysis of natural data beyond the sentence boundary; analysis of language use in society; and analysis of the properties of interaction between speakers/writers in everyday communication. Therefore, the distinction between discourse and text is quite blurred. This point also resonates in Crystal (1995), who says the distinction is not clear cut; both can be broadly used to include all language units with a definable communicative function, whether spoken or written. It is in this sense that the terms are used here, synonymously. We need to further distinguish these from register, the other interest of this study. Now, each text is resident in a domain, is of a type. The typology of texts/discourses is register, particular kinds of texts. In Biber (1999), this distinction is made explicit; register is used for situational varieties of English, e.g. newspaper language, academic language and the like, and each extended sample of language from a register is text. He notes that even a casual look at texts from different registers reveals extensive linguistic differences. This observation strengthens this effort, the examination of different text types or registers to see whether the use of multi-word items serves as an index of variability. But first, let us examine the literature of multi-word items in order to situate the study.

### A. Multi-word Items: A Brief Review

The concept of multi-word items in language is foregrounded by the idea that language occurs not as isolated units, but in chunks and groups, that language consists of semi-preconstructed units or prefabricated units which users utilize as single choices. This is the stance of some corpus linguists, like Sinclair (1991), and others, like Lewis (1997), who propose a lexical approach to language study. Lewis (1997, p.7) argues that 'language consists of chunks which, when combined, produce continuous text. The chunks are of different kinds, four of which are identified. One of these consists of single words while all the others are multi-word items'. Since the multi-word item is the major thrust of this study, it is prudent to define it, identify its types, and describe their characteristics. It is also important to point out that chunks in language are given different labels, as will be demonstrated shortly, though their functions and characteristics remain the same. In this study however, we shall adopt the term 'multi-word items', as the term is defined in the following section.

### B. Definition

It is necessary, from the outset, to point out that a definitional problem attends the term 'multi-word items', as signalled by its varying nomenclatures. To some scholars, it is a lexical phrase, for example Nattinger and DeCarrico (1992); to some it is formulaic language, for example (McCarthy, 1990); to some it is prefabricated language, for example Pawley and Syder (1983); to some it is multiword units, for example Lewis (1993, 1997), and to Hsu (2006), multi-word lexical units. To some others it is fixed expressions, for example Cowie (1988); to Sinclair (1991, 1998) it is idioms or pre-constructed language; to Moon (1997), it is multi-word items; and to Biber (1999), it is lexical bundles. Although all the scholars above make references to this phenomenon, it is Moon (1997, p. 43) who attempts a full definition of this concept, one which this study shall adopt. She states that:

...a multi-word item is a vocabulary item which consists of a sequence of two or more words ... This sequence of words semantically and/or syntactically forms a meaningful and inseparable unit. Multi-word items are the result of lexical (and semantic) processes of fossilisation and word-formation, rather than the results of the operation of grammatical rules.

The other definition of multi-word items worth noting is that of Nattinger and DeCarrico (1992, p. 37):

...prefabricated phrases are collocations if they are chunked sets of lexical items with no particular pragmatic functions; they are lexical phrases if they have such pragmatic function.

In this definition, they make a distinction between two types of multi-word item: collocations and lexical phrases, as does Lewis (1993, p. 92) when he says that '...looking at groups of lexical items is revealing. The most important groups are collocations, which are message-oriented, and institutionalized expressions, which are essentially pragmatic in character'.

Obviously, the terminological differences inherent in the concept of multi-word items are now evident. What this portends for lexical analysis is that each researcher needs to define carefully the terms, as they apply to their study. In this study, a multi-word item is defined, following Moon, as a vocabulary item that consists of a sequence of two or more words which semantically or syntactically form a meaningful and inseparable unit. In fact, Moon explains that multi-word items are very closely related to collocations, such that they can be seen as extreme cases of collocation. This was also the observation of McEnery, Xiao and Tono (2006), who note that multi-word items are closely associated with collocations and have been an important topic in lexical studies.

However, Nattinger and DeCarrico (1992, p. 36) distinguish between multi-word items (which they call lexical phrases) and collocations. According to them, 'collocations are strings of specific lexical items, such as *rancid butter* ... that co-occur with a mutual expectancy greater than chance. These strings have not been assigned particular pragmatic functions'. In contrast, multi-word items 'are collocations that have been assigned pragmatic functions, and they are of two main types; strings of specific (non-productive) lexical items, which allow no paradigmatic or syntactic substitution (and) generalized (productive) frames, which have been assigned a pragmatic function'. Therefore, phrasal combinations in English are collocations if they are chunked sets of lexical items with no pragmatic functions, but they are multi-word items if they have pragmatic functions (Nattinger & DeCarrico). They note, as Moon does, that variation exists between multi-word items but that their categories form a continuum rather than sharp boundaries. The four structural criteria which characterize them, they say, are length and grammatical status, canonical or non-canonical shape, variableness or fixedness, and continuous or discontinuous nature. They note the four main types as polywords, institutionalized expressions, phrasal constraints and sentence builders, all of which differ in their structure and functions to varying degrees. Concerning types, Moon says that multi-word item is a super-ordinate term for compounds, phrasal verbs, idioms, fixed phrases, and prefabs. Another class, mentioned by Tschichold (2000), is the noun phrase.

In yet another attempt, Biber (1999) explains the concept of multi-word items. According to him, many multi-word expressions function in English as structural or semantic units, and the most common of these are phrasal verbs and prepositional verbs. And these constitute single entities with respect to both their structure and meaning. He further notes that these are of different kinds depending on their degrees of idiomaticity and invariability. At one extreme are idioms, invariable expressions whose meanings are not a sum of their parts; then there are collocations which are lexical word associations which co-occur more frequently than could be expected by chance; and at the other extreme are

lexical bundles, word sequences (normally of two or more words) which show a tendency to recur in natural discourse, irrespective of idiomaticity and structural status.

Moon (1997), in addition, explains very clearly how to establish a sequence of words as a multi-word item. First, is the factor of institutionalization which is the degree to which a multi-word item is conventionalized in the language. Second, is fixedness, which is the degree to which a multi-word item is frozen as a sequence of words. And third, is non-compositionality, which is the degree to which a multi-word item cannot be interpreted on a word-by-word basis, but has a specialised, unitary meaning. These, she adds, are however, not absolutes but variables, and are present in differing degrees in each multi-word item. We shall examine, albeit briefly, the different types of multi-word items in order to clarify our discussion of them in subsequent sections of the study. To illustrate them, we shall rely on the accounts by Lewis (1993, 1997), Nattinger (1980) and Hsu (2006), shown below.

### C. Types of Multi-word Items

Lewis (1993, 1997) identifies different types of multi-word units (Lewis' cover term) as follows:

Multiword items: He identifies the different types as polywords, e.g. *by the way*, arbitrary combinations, e.g. *to and fro*.

Collocations: Their types were identified as fixed and free, e.g. *blonde hair*, *strong tea*.

Institutionalized Expressions: These are pragmatic in nature. For example, *not yet*, *just a moment*, *please*, etc.

Sentence heads or frames, e.g. *Sorry to interrupt... but can I just say... .*

Fixed Expressions: Their types are social greetings, e.g. *good morning*, politeness phrases, e.g. *no thank you*, idioms, e.g. *do not make a mountain out of a mole hill*.

Semi-fixed Expressions: Most expressions fall into this category. Lewis exemplifies with spoken sentences with a simple slot, as in *Could you please pass the... .*, sentence heads which can be completed in variety of ways, e.g. *The football match was interesting/ boring/ annoying*.

In another study, Nattinger (1980) classifies multi-word items (which he terms lexical phrases) into six groups as follows:

1. Polywords: short fixed phrases, whose meaning is often not analysable by the regular rules of syntax, and which can substitute for single words, e.g. *kick the bucket*.

2. Phrasal constraints: short, relatively fixed phrases with slots that permit some variation, e.g. *a ... ago*.

3. Deictic locutions: short- to medium-length phrases of low variability used to monitor conversation, e.g. *as far as I know*.

4. Sentence builders: phrases of up to sentence length – highly variable phrases containing slots, e.g. *not only X but Y*.

5. Situational utterances: usually complete sentences, e.g. *I'll see you next week*.

6. Verbatim texts: e.g. numbers, the alphabet, days of the week, aphorisms and proverbs.

Hsu (2006), in his investigation of multi-word units in ELT textbooks, uses the term multi-word lexical phrases and identifies four types as: lexical collocations, fixed expressions, semi-fixed expressions, and idioms; and describes each group thus:

Lexical collocations: Word associations where one word recurrently co-occurs with one or more other words as the only, or one of few, possible lexical choices, e.g. *blonde hair*.

Fixed expressions: Word groups which allow no replacement in any of their components, e.g. *what's up?*

Semi-fixed expressions: These are similar to fixed expressions except they allow replacement in at least one or more of their components, e.g. *If I were you, I would... .*

Idioms: Word groups fixed in word order and substitutability, e.g. *a skeleton in the closet*.

We must point out, however, that the multi-word item in English is not as productive in its kinds as these classifications seem to suggest. Rather, each writer/researcher distinguishes them as suits his or her purpose. Hence, their typologies, at present, are not mutually exclusive. This, in itself, is indicative that this area of study is still at an elementary stage because, as yet, there are no established terminologies, classifications and boundaries for this linguistic phenomenon. In the present study, an integrated classification is made of seven main types; compounds, phrasal/prepositional verbs, fixed expressions, semi-fixed expressions, noun phrases, lexical collocations, and idioms. The interest this phenomenon has generated so far in language research and, most importantly, its significance in language variation, are the foci of the next section.

### D. Multi-word Items and Register Differentiation

Much of the work done on the multi-word item is based on its role in language learning. It is argued that mastery of multi-word items in a target language enhances the communicative competence of the second language learner, and thus enables him or her to produce appropriate sentences in the language (Lewis, 1997; Stubbs, 1996; Nattinger & DeCarrico, 1992). According to Nattinger and DeCarrico (p.1), 'current research converges in a way that reveals the lexical phrase as an ideal unit which can be exploited for language teaching'. This is based on the pervasive role that ritualization plays in language behaviour; routinized formulas and other forms of prefabricated language chunks, products of this ritualization, seem to play a large part in both acquiring and performing language (Nattinger & DeCarrico). Therefore, beyond playing a role in language description, they see a possible connection between multi-word items and language theories of competence, performance and pragmatics. Moon (1997) is in no less agreement on the significance of multi-

word items in language learning when she remarks that they pose a problem to the second language learner. Many other works deal with their function in the complete description of the lexis, even of particular genres (Nelson, 2000; Sinclair, 1991). In Nelson, for instance, a thorough description of the lexical network of business is given with the identification of business-specific multi-word items playing a crucial role in painting a picture of business English, and subsequently in postulating its extended function of enhancing business-related communications and learning.

Only a few other works have enquired into the role of multi-word items in language variation to establish multi-word items as an index of text variability. For example, Biber (1999) found that they are a reliable indicator of register variation. He showed that the structural types and distribution of multi-word items differ across text types of conversation, classroom teaching, textbooks, and academic prose. Earlier, Moon (1997) has observed that multi-word items vary according to regional varieties of English, and that the lexical densities of multi-word items in particular text types prove that there are strong genre-specific preferences, e.g. idioms are associated with journalism. It is to this category of work that this research seeks to make a contribution.

Some other works, for example Macaulay (2003), note that the complexity of discourse-variation study – finding out what patterns of language characterize language use in particular settings – requires many methodological decisions not crucial to studying other kinds of variation. He notes, for instance, that discourse variation can best be said to be only at an elementary stage because ‘there is no agreed methods of collecting, or analyzing data, on what features are suitable for investigation, on how to identify possible discourse features, and what significance to attach to the use of a particular feature’ (Macaulay, 2003, p. 284). Concerning earlier findings on discourse variation, he remarks that there are a number of intriguing claims that need to be tested again and again, by the same or different methods, in similar or different settings, with similar or different samples, noting that it is only when there is convergence of results that confident claims about discourse differences are possible. Against this backdrop, then, this study gains considerable significance because it will serve both to retest previous results by its choice of the academic and fictional writings previously examined by Biber (1999), and to make fresh findings by examining the character of multi-word items in political, religious and memorial speeches, whose text samples will be presented shortly.

## II. DATA AND METHOD OF THE STUDY

As indicated above, the data for this study are composed of extracts from five types of text: academic and fictional writings, and political, religious and memorial speeches. The choice of these registers is informed by their extensive use. Except for memorial speeches, language use for fiction, academic purposes, religion and politics is certainly very wide. Our sample texts were extracted randomly from these domains. It is imperative to point out that only short excerpts of them can be accommodated in a work of this nature. The method that will be used here is both qualitative and intuitive. This entails a comparative textual analysis of this linguistic item across the text types to investigate similarities and dissimilarities of its use. The overall objective, of course, is to examine the use of multi-word items in these contexts to see whether the texts can be differentiated on the basis of their use of this item. If yes, then the results become relevant to the learning and use of language in these contexts.

In line with the objectives of the study, we shall first identify and discuss instances of the use of formulaic language (multi-word items) in each of the texts. These multi-word items shall be underlined for recognition. Then, second, we shall compare the types found across the text types to see whether particular registers show preference for certain multi-word items above other types. In other words, we will investigate if multi-word items characterize specific registers and, in doing so, confirm or disprove the existence of genre-specific multi-word items as posited by some previous research, e.g. Biber (1999). We shall now present and analyse each data type below, with overall conclusions following immediately.

## III. DATA PRESENTATION AND ANALYSES

### *Text Sample 1: Fictional prose writing*

Jess? Her mother’s voice sounded through the hallway, mixing with the mustiness around her so well that the sound almost had a smell. To Jess, sitting in the cupboard, the sound of her name was strange, wobbly, misinformed, as if she were inside a bottle, or a glass cube, maybe, and mum was outside it, tapping. I must have been in here too long –

‘Jessamy’ Her mother’s voice was very stern. Jessamy Harrison did not reply. She was sitting inside the cupboard on the landing, where the towels and other linens were kept, saying quietly to herself, I am in the cupboard. She reminded herself that she was in the cupboard, so that she would know exactly where she was, something that was increasingly difficult each day. Jess found it easier not to remember, for example, that the cupboard she had hidden in was inside a detached house, on Langtree avenue. (Oyeyemi, Helen. The Icarus Girl 2005, p.1)

### *Analysis*

In this text, it can be seen from the underlined words that almost all the multi-word items found are compounds. There is, however, more use of noun compounds, for example *hallway* and *cupboard* than other compounds, for example adjective + noun compounds which occurred once, e.g. *glass cube*. Other multi-word items used are fixed expressions, such as *for example*, and noun phrases, e.g. *Langtree avenue* and *a detached house*. However, the tendency is to speculate a preference for compound words in fictional prose due to their frequent occurrence.

*Text Sample 2: Academic writing*

So remarkable, then, is the act of writing that it apparently frees some part of us from the confines of calendar and atlas. Through it we reach to other times and other places; it passes the word to Whitsuntide and fetches an answer from Athens or Arkansas. Yet this, the wonderful liberty of the medium, is hardly achieved at a stroke. We begin by talking to ourselves, convincingly enough, maybe, but presently become involved in a long and laborious process of translation, a dull and often slavish imposition on the spirit. Suddenly language itself is strange, intractable, a material that responds lumpishly, a substance of stubborn grain that must be worked deliberately and in silence. Here is a speechless acquisition, a clenched and staring discipline. Here, indeed, is a trade, like metal work or joinery; through practice we learn the craft, through learning we extend the practice, and always there is that charmed procedural circle, of doing and learning and doing, into which the initiate must break unaided. Demonstration is helpful, analysis is doubtless of great value, but who can teach the synthetic power of the creative imagination as it buds forth some artefact – say a table, say a text? There are bounds, of course, to what theorizing can do; and so perhaps we have come in this book to a conjectural limit, beyond which all attempts at advice might be deemed presumptuous and futile. Once the making has begun, the maker must be his own apprentice and his master. (Nash, W. Designs in Prose 1980, p. 159)

*Analysis*

In this academic prose text, we find higher productivity in the use of multi-word items than in the prose fiction text. For instance, there is the use of compounds: adjective + noun compound, e.g. *creative imagination*, *conjectural limit*; prepositional verbs, e.g. *reach to*, *involved in*, *come to*; fixed phrases, e.g. *at a stroke*, *of course*; noun phrases, e.g. *a speechless acquisition*; and idioms, e.g. *the confines of calendar and atlas*. There is quite a number of idioms in this text and that is, to say the least, surprising for an academic text. Also, there is an equally good number of prepositional verbs. This leads us, again, to attempt prediction of the text's preference for idioms and prepositional verbs.

*Text Sample 3: Religious Speech*

Nowhere is the tragic tendency to conform more evident than in the church, an institution which has often served to crystallize, conserve, and even bless the patterns of majority opinion. The erstwhile sanction by the church of slavery, racial segregation, war, and economic exploitation is testimony to the fact that the church has harkened more to the authority of the world than to the authority of God. Called to be the moral guardian of the community, the church at times has preserved that which is immoral and unethical. Called to combat social ills, it has remained silent behind stained-glass windows. Called to lead men on the highway of brotherhood and to summon them to rise above the narrow confines of race and class, it has enunciated and practiced racial exclusiveness ... We have become showmen to please the whims and caprices of the people. We preach comforting sermons and avoid saying anything from our pulpit which might disturb the respectable views of the comfortable members of our congregations. Have we ministers of Jesus Christ sacrificed truth on the altar of self-interest, and like Pilate, yielded our convictions to the demands of the crowd? (King, M. L. Strength to Love 1963, p. 21)

*Analysis*

In the religious text we see, too, instances of compound words: noun compounds, e.g. *majority opinion*, *showmen*; adjective + noun compound, e.g. *moral guardian*; and verbal + compound, e.g. *comforting sermons*. Other multi-word items found are: prepositional verbs, e.g. *serve to*; phrasal verbs, e.g. *rise above*; fixed phrases, e.g. *whims and caprices*; idioms, e.g. *sacrifice truth on the altar of self interest*; and noun phrases, e.g. *ministers of Jesus Christ*. Looking at the identified word sequences, it is clear that compounds prevail and this implies an inclination toward their use in religious texts.

*Text Sample 4: Political Address*

OBAMA: My fellow citizens:

Forty-four Americans have now taken the presidential oath. The words have been spoken during rising tides of prosperity and the still waters of peace. Yet, every so often the oath is taken amidst gathering clouds and raging storms. At these moments, America has carried on not simply because of the skill or vision of those in high office, but because we the people have remained faithful to the ideals of our forebears, and true to our founding documents.

So it has been. So it must be with this generation of Americans.

That we are in the midst of crisis is now well understood. Our nation is at war, against a far-reaching network of violence and hatred. Our economy is badly weakened, a consequence of greed and irresponsibility on the part of some, but also our collective failure to make hard choices and prepare the nation for a new age. Homes have been lost; jobs shed; businesses shuttered. Our health care is too costly; our schools fail too many; and each day brings further evidence that the ways we use energy strengthen our adversaries and threaten our planet. (Excerpt from text of President Barack Obama's Inaugural Address: p.1)

*Analysis*

The political speech makes use of such multi-word items as: compounds, e.g. *fellow citizens*, *founding documents*, *health care* (all noun compounds); prepositional verbs, e.g. *carried on*; phrasal constraints/semi-fixed expressions, e.g. *a new age*; noun phrases, e.g. *the presidential oath*; and idiomatic expressions, such as *still waters*, *gathering clouds*, *rising tides*, *raging storms* and *hard choices*. There seems to be a greater use of idioms in the text which is suggestive of preference.

*Text Sample 5: Memorial Address*

My President, the Senate has assembled today to remark for the record the death of John Fitzgerald Kennedy, President of the United States. I shall be brief, for his life, too short, shut off too soon, speaks for him. In these last hours, a profile in courage has emerged from the emulsion of his death, and did not experience of the Nation and the world.

John Fitzgerald Kennedy's courage was the human courage, the courage which all must have merely to live in this world, in the ever-present shadow of death, it was the special courage to defy the cold hand of death when it reaches out too eagerly, and as twice it did in the wounds of the war and in the grave illness of his senate years. It was the quiet courage to accept death's finality when it would be denied no longer.

This is the profile of the man who walked among us not long ago on the floor of the Senate. This is the profile of the man who emerged to reawaken the Nation to its finest meaning. This is the man who struck new sparks of hope in a world dark with unspeakable fears. (Senator Mansfield's speech on the death of J. F. Kennedy)

#### Analysis

The memorial speech contains a number of multi-word items. There are a good number of idioms such as *last hours*, *cold hand of death*, *the wounds of war* and *a dark world* which give an idea of the field of discourse. Then there are phrasal and prepositional verbs, for example *shut off*, *reach out* and *walked among*. We have, too, uses of fixed expressions, e.g. *for the record*, noun phrases e.g. *the floor of the senate*, and compounds, e.g. *human courage*, *special courage*. Here, too, idioms are seen to be more prevalent.

Before discussing the findings on the multi-word items found across the texts, it is important to classify the results in order to show at a glance what types characterize each text. The discussion, which will follow, shows to what extent these multi-word types can be said to distinguish each text type or register. Table 1, below, shows the distribution of the types of multi-word items found in the texts.

TABLE 1  
TYPES OF MULTI-WORD ITEMS FOUND IN THE TEXTS

| Text Types       | Comp. words | Phrasal Verbs | Prep. Verbs | Phrasal Const. | Noun phrase | Fixed Exp. | Idioms |
|------------------|-------------|---------------|-------------|----------------|-------------|------------|--------|
| Fictional Prose  | •           |               |             |                | •           | •          |        |
| Academic Prose   | •           |               | •           | •              | •           | •          | •      |
| Religious Speech | •           | •             | •           |                | •           | •          | •      |
| Political Speech | •           |               | •           | •              | •           | •          | •      |
| Memorial Speech  | •           | •             | •           |                | •           | •          | •      |

#### IV. DISCUSSION OF FINDINGS

We find that, except for the fictional text, the other texts demonstrate a fair distribution of multi-word types, and that while compounds, prepositional verbs, noun phrases, fixed expressions and idioms are common to the other four texts, phrasal verbs and phrasal constraints are relatively rare. They each occurred in two out of the five texts. The scant use of multi-word items in the fictional prose text is strange. This is especially so because previous study, e.g. Biber (1999), establishes multi-word items, such as phrasal verbs and prepositional verbs, as being particularly frequent in fiction because they are colloquial in tone and so amenable to such constructions as phrasal verbs. Although the point stands to reason, the present study does not corroborate it, due probably to the length of the text.

As can be seen, compounds were found across the text types which indicate that their use is not distinguishing in itself, except for its types. This means that all the texts are expected to manifest similarity of discourse content, since compounds perform the discourse function of denoting and carrying high information content. Nonetheless, each text demonstrates preference for certain structural types of compounds. For example, in fiction, there were more noun compounds, in academic prose, there were more Adj. + N compounds, in the religious text, there were mostly Adj. + N compounds, in the political speech text, more noun compounds were found, and in the memorial address, both noun compounds and Adj. + N compounds were seen to have equal occurrences. Each use relates to the discourse structure of the particular text.

The observation of high information content is further supported by the common occurrence of noun phrases in all the texts. As is well known, nouns are the main carriers of lexical meaning as they make reference and convey much information. The core of a noun phrase however points to words key to each text type and, to that extent, can be said to distinguish the texts. For example, in the political text, we find such key phrases and words as *presidential oath*, *high office*, *fellow citizens*, *our nation*, *vision* and *forebears*, which point us toward political discourse.

Also found common to all the texts are fixed expressions which, as Moon (1997) observes, organise text and provide a framework for an utterance or the argument in a text. The fixed expressions found perform exactly that discourse function, here in the texts. Take for example, the use of *of course* in the academic text, and *whims and caprices* in the religious text; both are structuring devices which serve as organiser in the case of *of course*, and summarizer in the case of *whims and caprices*.

Noteworthy too is the occurrence of idioms in four out of the five text types. According to Moon, idioms typically evaluate, connote and convey judgements, but also function in discourses as prefaces or summarizers. We see, in each instance, idioms in the texts functioning in one of these ways. But beyond saying that these text types utilize idioms either to a greater or lesser extent, we are not confident enough to hypothesize that they characterize a text type. However, we are content to say that political language and the language of memorial speech demonstrate a greater tendency to idiomaticity than academic prose, religious speech or, especially, fictional prose. The reason could be that political and memorial discourses require more evaluation, judgement and suggestion than those others which need to denote and convey information plainly.

These analyses show, in addition, that the broad categorisation of language into speech and writing does not imply that both are markedly differentiated in their linguistic characters, rather we find areas of both convergence and divergence. This is illustrated by the results which demonstrate the closeness of behaviour of multi-word items in political speech and, for example, academic prose. Biber (1988) deals extensively with this in his phenomenal publication, *Variation across Speech and Writing*.

## V. CONCLUSION

Multi-word sequences are proved to be universal in language and thus need to be acquired by the language user. This poses a greater problem to the second language learner who has to grapple not only with the single lexis of the second language, but also combined sequences of words which are considered difficult either to teach or learn. This difficulty is due to their non-compositionality and pragmatic nature. But, owing to their importance for the competent use of language, especially to the L2 speaker to whom they are a measure of proficiency, the knowledge and use of multi-word items in appropriate situations can hardly be exaggerated. This was the point of Biber (1999), who says that it is important for the foreign language learner of English to know that learning to produce natural idiomatic English is not only a matter of making well-formed sentences, but also of using well-trieved lexical expressions in the appropriate places. This is because lexical expressions are language building blocks used by English speakers and writers in different situations. Therefore, their mastery facilitates language use in different contexts. As a result, their examination in different genres, to establish both their use and variation, is a desideratum in language research, especially in light of the variableness of language.

In examining this phenomenon in the context of different registers of English, this study draws attention to it, so as to encourage its acquisition such that communication in various genres can become more proficient. This leads us back to the question at issue – are multi-word items distinguished according to registers? In response to this question, the findings of this research provide the following answers:

1. The study reveals the existence of multi-word items in all five registers examined. They occurred to varying degrees, for example, the fictional text manifested the least productivity, while the academic text showed the highest productivity in the use of multi-word sequences. This leads to the conclusion that some registers are more formulaic than others, which in turn implies that learning to use language in domains where multi-word sequences are frequent may be more demanding. To this extent, we can say that texts can be distinguished according to the density of multi-word items in them, as Moon (1997) also found.

2. The results acknowledge that particular registers show preference for certain types of multi-word items rather than others. For instance, although all five registers utilized compounds, we found that while fiction used mostly noun compounds, others, like academic prose and religious speech, preferred adjective + noun compounds. Beyond selection of sub-types of multi-word items, there is evidence of choice of whole types of multi-word items. For example, we saw that the language of politics and memorial addresses were more idiomatic than the others. And this choice was seen to impact on the discourse structure of the texts. To this end, also, we recognize register differences.

3. But the study did not fully concur with the results of former work, especially with regard to which multi-word items are known to occur in which registers. For example, Biber's (1999) claim that phrasal verbs and prepositional verbs are frequently used in fiction was not supported by the present research. This means that other factors are at play in the determination of what multi-word items a writer employs or does not employ. Or, as Macaulay (2003) observes, we are yet to arrive at a convergence of results on the phenomenon.

4. Finally, there is some indication that the multi-word item can serve as an index of text variability, but precisely which multi-word items are expected to occur in a register, or not occur, remains inconclusive. What is also noteworthy is that beyond text type, other indices of choice exist which determine the choices of the language user. At best, then, the role of multi-word items in register variation remains a topic of on-going research.

We therefore recommend further and continual research into the use of this item until a complete picture emerges. And such a picture can only emerge reliably through corpus investigations. Corpora, it is noted, provide ample language evidence and a realistic foundation for the study of language, particularly the comparison of language varieties (Svartvik, 1996). And as Sinclair (2004) says, we should trust the text, be open to what it may tell us, and rebuild our picture of language and meaning.

## REFERENCES

- [1] Baker, P. (2006). Using corpora in discourse analysis. London: Continuum.
- [2] Biber, D. (1988). Variation across speech and writing. Cambridge: Cambridge University Press.
- [3] Biber, D. (1999). Lexical bundles. In Biber, et al. (eds.), *Longman grammar of spoken and written English*. London: Longman.
- [4] Barrack, O. (2009). President Barack Obama's Inaugural address. [www.time.com/time/politics/article/0,8599,1872715,00.html](http://www.time.com/time/politics/article/0,8599,1872715,00.html) (Accessed 20/01/ 2009).
- [5] Cowie, A. P. (1988). Stable and creative aspects of vocabulary use. In Carter, R. & McCarthy, M. (eds.), *Vocabulary and language teaching*. Harlow: Longman.
- [6] Crystal, D. (1995). The Cambridge encyclopaedia of the English language. Cambridge: Cambridge University Press.
- [7] Halliday, M A. K. & Ruquaiya Hasan (1976). Cohesion in English. London: Longman.
- [8] Halliday, M. A. K. & Hasan, R. (1989). Language, context, and text. Oxford: Oxford University Press.
- [9] Hsu, J. (2006). An Analysis of the multiword lexical units in Contemporary ELT Textbooks. Paper presented at the International Conference on English Teaching and Learning in the Republic of China (23rd, Kaohsiung, Taiwan, May 27-28, 2006).
- [10] King, M. L. (1963). Strength to love. London: William Collisons.
- [11] Lewis, M. (1993). The Lexical approach. Hove: Language Teaching Publications.
- [12] Lewis, M. (1997). Implementing the lexical approach. Hove: Language Teaching Publications.
- [13] Macaulay, R. (2003). Discourse variation. In Chambers, J. K., Trudgill, P. & Schilling- Estes, N. (eds.), *The handbook of language variation and change*. Wiley: Blackwell.
- [14] McCarthy, M. (1990). Vocabulary. Oxford: Oxford University Press.
- [15] Moon, R. (1997). Vocabulary connections: multi-word items in English. In Schmitt, N. & McCarthy, M. (eds.), *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- [16] McEnery, T., Xiao, R. & Tono, K. (2006). Corpus-based language studies: an advanced resource book. New York: Routledge.
- [17] Nash, W. (1980). Designs in prose. London and New York: Longman.
- [18] Nattinger, J. R. (1980). A lexical phrase grammar for ESL. In *TESOL Quarterly* 14, 337-44.
- [19] Nattinger, J. R. & DeCarrico, J. S. (1992). Lexical phrases and language teaching. Oxford: Oxford University Press.
- [20] Nelson, M. (2000). A corpus based study of business English and business English teaching materials. Ph.D Dissertation, University of Manchester.
- [21] Oyeyemi, H. (2005). The Icarus girl. London: Bloomsbury.
- [22] Pawley, A. and Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Richards & Schmidt (eds.), *Language as communication*. Harlow: Longman.
- [23] Senate Document No. 59 (1964). J. F. Kennedy's Memorial addresses. Washington: United States Government Printing Office.
- [24] Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.
- [25] Sinclair, J. (1998). The lexical Item. In Weigand (ed.), *Contrastive lexical semantics*. Amsterdam: John Benjamins Publishing Company.
- [26] Sinclair, J. (2004). Trust the text: language, corpus and discourse. London: Routledge.
- [27] Stubbs, M. (1996). Text and corpus analysis. Oxford: Blackwell.
- [28] Svartvik, J. (1996). Corpora are becoming mainstream. In Thomas, J. & Short, M. (eds.), *Using corpora for language research*. London and New York: Longman.
- [29] Tschichold, C. (2000). Multi-word units in natural language processing. Hildesheim: George Olms Verlag.



**Alexandra Uzoaku Esimaje** was born in Oguta, Nigeria, in August, 1962. Esimaje was formally educated in the country and obtained a first school leaving certificate from Shell Camp Primary School, Owerri, in 1972, a West African School Leaving Certificate from Owerri Girls Secondary School in 1979, a Bachelor of Arts Degree (B.A. English) from University of Ilorin in 1984, a Master of Arts Degree (M.A. English) from University of Maiduguri in 2000, and in University of Maiduguri, has just completed a Doctor of Philosophy Degree (P. h D. English) in 2011.

She has worked as teacher and lecturer in various institutions of learning for a period of over twenty years. Presently, she lectures English in Benson Idahosa University, Benin City, Nigeria, where she was the Head of the English Department and has also held several other positions. She is widely published nationally and has some foreign publications to her credit. She has authored one book titled *Oral English and Grammar*. Jos: Lecaps Publishers, (2002).

Dr. (Mrs.) Esimaje is member of many professional bodies such as Nigeria English Studies Association (NESA), English Language Teaching Today (ELTT) and Nigeria Association for Semiotics Studies (NASS). She holds the fellowship of the American Council of Learned Societies (F.ACLS) having won the 2010 African Humanities Programme award. She has also worked for the British Council as instructor at a year-long national workshop for training teachers of English in Nigeria. For four years now, she has been involved in a similar project by the Nigerian Government to re-train teachers of English in Primary Schools.