Deverbal Nominalisations across Written-spoken Dichotomy in the Language of Science

Mohammad Hussein Norouzi Faculty of Foreign Languages and Literatures, University of Tehran, Iran Email: mhnorouzi@ut.ac.ir

Ali Akbar Khomeijani Farahani

Department of English Language and Literature, Faculty of Foreign Languages and Literatures, the University of Tehran, Iran

Email: farahani@ut.ac.ir

Davood Borzabadi Farahani

Department of English Language and Literature, Faculty of Foreign Languages and Literatures, the University of Tehran, Iran

Email: borzabad@ut.ac.ir

Abstract—Scientific language has always been both a source of difficulty for the students of science and, in the recent decades in particular, a theme for speculation for the scholars seeking to describe and analyse language as it is used and conceived. The language of science has always been presumed to be more difficult than the language used in other fields. Scientific concepts and their sophistication are one source of its difficulty. The writing style and dynamics preferred by writers in the fields of science is commonly considered another source of such difficulty. Nominalisation, the use of derivative nouns of verbs and adjectives, which is more prevalent in the language of science compared with the language used elsewhere, is another feature attested to by many scholars. The present study aimed at comparative investigation of the phenomenon of deverbal nominalisations (nominalisations with verb origin) across written and spoken scientific language. The spoken samples were chosen from BBC's programs in the genre of science and the written samples from science textbooks and credited science magazines (different issues). The samples in each modality included 400,000 words (for each modality) amounting to 800,000 words as the total size of the corpora considered. The findings showed a higher frequency of verbs in spoken modality and of nominalisations in written. Another interesting point was the highest frequency of material process type in both verbs and deverbal nominalisations. (The typology of processes from the Hallidayean systemic functional grammar framework was the basis for classification of verbs and their respective nominalised editions.) Relational process type which claimed the second highest frequency among verbs showed a much lower frequency among nominalisations. The consideration of the slots accommodating the nominalisations in the sentences also revealed insightful points about possible differences between spoken and written language (of science) in terms of the concentration of semantic and lexical density.

Index Terms—systemic functional grammar/linguistics, grammatical metaphor, nominalisation, spoken scientific language, written scientific language, process types

I. INTRODUCTION

Scientific language, the language used to convey scientific concepts and notions, has attracted the attention of scholars in the field of applied linguistics for the purposes of both understanding this specific type of language and of helping the students of science acquire the specific register of language with functional levels of success. It has prevalently been associated with the reputation of being more difficult and complex than non-scientific registers. Beside the complexity of scientific notions, which are normally beyond the daily and routine demand of knowledge requirements, the diction and method scientists opt to employ in the disciplinary corpora is unanimously agreed upon as a main source of complexity. The authors of scientific corpora tend to avoid first person singular case and likewise try to avoid expression of ideas as their personal beliefs or understandings and, therefore, incline to use projections such as 'it can be claimed/concluded that' instead of adverbs like 'seemingly' or 'I presume that...'. This serves the purpose of making the arguments appear more impersonal, and thus, acceptable, meanwhile being another source of complexity in scientific discourse.

One of the major ways for creating such complexity is the attempt to include more information in a smaller number of words than it is common in normal language. An author can subsume a whole proposition, which is normally expressed via a stand-alone clause, in another clause, encompassing more than a single proposition in a single clause. This serves manifold purposes. It reduces the number of words in favour of a higher level of lexical and semantic density. Moreover, it objectifies the act taking place in the original proposition and can either justify it by indirectly taking it for granted, or removing the responsibility of the agent through its omission from the clause. Based on the content and context of sentences, such a strategy can serve different purposes. In this process, the verb of the original proposition loses its agent and the nominalised version of the verb is employed, which is void of agent or time, and is considered an existent reality, or in Fairclough's (2003) terms, it is objectified. Moreover, it functions as a satellite for another verb, adding more semantic load to the typical scope hinging around any single verb.

II. BACKGROUND

A. Metaphor

Metaphor is, loosely put, the use of a lexical form for a semantic concept different from the semantic concept it is primarily intended for. In a more extensive use, it is the utilisation of a conceptual device for a different purpose than its originally intended one. In order to elucidate a concept, for instance, people might resort to a different but similar or analogous concept to convey their intentions. Therefore, it is a means of conveying and developing the realm and depth of knowledge. Baake (2003), in a similar token, believes that different types of metaphor have a role in the production, development or fine-tuning of knowledge. Science, according to him, "has never been without metaphor and related figures of thought and speech" (p. 68). He mentions Copernicus's metaphor of sun as a lantern on a hill as an illustrative example from the early days of modern science.

B. The Notion of Metaphor and Its Grammatical Edition

The ability to utilise a form to a (semantic) content different from, and in most cases similar to, that originally intended for that specific form constitutes the basis for literally all metaphoric use of language. Such ability plays a crucial role in any human being's introduction to and accommodation in the system of human existing and developing knowledge. Metaphor is considered one of the concepts linking semantics and lexicogrammar and also a good tool to study the semogenesis, or the evolution of meaning in cultures and individuals (Steiner, 2002-2003, *inter alia*).

Brought to the fore mainly through the works of the prominent systemicist, M.A.K. Halliday and his followers, grammatical metaphor enjoys a similar pivotal position in the development of science and expounding its various areas. In grammatical metaphor, it is the grammatical category (parts of speech or word classes) that falls beyond the common expectation. In sum, lexical metaphor includes the change of meaning while the word class remain the same (a noun taking up a different ideational meaning), while in the case of grammatical metaphor, the ideational meaning is consistent while the word class used to express it changes (e.g. *implement* and *implementation* have the same basic concept of a process) (Halliday, 1994; Halliday and Matthiessen, 1999); this is called transcategorisation. According to this notion of grammatical metaphor, first, every element in any language belongs in a category or class of words, and second, some words can depart from the word class they originally belong in. Therefore, semantic entities would likewise be expressed through a word class different from the originally intended one (Halliday and Matthiessen, 1999) and 2004).

Systemic functional linguistics considers three main metafunctions for language, which are realised in any piece of language use: *ideational, interpersonal* and *textual*. The ideational metafunction includes our understanding of the world and how notions and entities are interrelated. The interpersonal metafunction substantiates the social role and potentials of language. Finally, the arrangement of and the emphasis on different parts and elements in utterances and clauses is what the textual metafunction deals with (Halliday and Hasan, 1976; Halliday, 1985 and 1994).

These three metafunctions of language are determined by three mainly parallel determinants. They are field, tenor, and mode. Parallel to the ideational metafunction is field, or the matter of content or what is being talked about. It is also referred to as discourse type. The second factor considers the people involved in a linguistic exchange and the social ranks and relations they have, which is in accordance with the interpersonal metafunction. Mode, the last factor, runs parallel to textual metafunction and pertains to the form and format of the text being produced and exchanged. A simple example of this last factor could be whether it is written or spoken (Halliday, 1994; Halliday and Matthiessen, 2004; Eggin, 2004; Thompson, 1996, 2004; Martin, Matthiessen and Painter, 1997).

The interplay between the closely connected sets of metafunctions and the context components (field, tenor and mode) confers on the language two types of choices to be made in each slot of a clause. They are paradigmatic and syntagmatic structures. *Paradigmatic structure*, or in better words *system*, is the source of power and potential for this theory of grammar, and also the source of the term *systemic* in the theory's name, that is, Systemic Functional Linguistics/Grammar. In paradigmatic system, in each stage of the production of a sentences or an utterance, the user has a full set of resources for each of the three main functions. Each choice made by the user is a choice denoting at least one of the three metafunctions. These three metafunctions should be considered simultaneously, or more appropriately, synchronously. Moreover, the choice of each item bears comprehensive implication of the possible (set of) choices to follow (hence *syntagmatic structure*) (Halliday, 2003; Matthiessen, Teruya and Lam, 2010). The result of the interaction between the two structures would be the unfolding of meaning on the part of the producer (speaker or writer) and its accumulation on the part of the receiving end (reader or speaker).

In systemic functional linguistics, within the ideational metafunction, each type of semantic notions or relations is *construed* through a particular type of linguistic items. Entities are normally construed through nouns, processes

through verbs and attributes through adjectives (Halliday, 1985 and 1994; Halliday and Matthiessen, 2004). Such oneon-one relationships are unmarked (or congruent in functional systemic terms) (Halliday, 1994; Thompson, 1996 and 2004, *inter alia*). Prepositions, however, or more comprehensively, functional words, are typically employed to express the relationships among the semantic notions conveyed through lexical items.

There are three types of grammatical metaphor corresponding to the three metafunctions. In the ideational grammatical metaphor, the unmarked respective relationship between the three semantic notion types of entities, processes and attributes and their related construal vehicles of nouns, verbs and adjectives is dislocated. In the most common form of grammatical metaphor, nouns are used to construe processes and attributes (*nominalisation*). Therefore, the relationship between the three basic semantic types and the construal vehicles is marked. In the interpersonal metaphor, utterances take up social roles different from their original ones. Finally, in textual metaphor, the theme-rheme structure deviates from the unmarked structure. It-cleft, wh-cleft and initial adverbial clauses and phrases are some examples of textual grammatical metaphor.

C. Processes

Processes are the denotations of the pivotal elements of semantic functioning of human cognition and in the lexicogrammar as well (Halliday and Matthiessen, 2004). Out of the three metafuctions, the ideational one, which treats language as representation of human experiences, is employed to convey the tenably broadest of the three fields of meaning served by language. This, however, is not intended to denigrate the other two metafunctions; it is simply wider and can accommodate a much more diverse range of what human beings can sense and conceive of, including constant flow of events like 'happening', 'doing', 'sensing', 'saying', 'being' or 'having' (Halliday and Matthiessen, 1999 and 2004). *Transitivity* is the system that governs the relationship among bits and pieces of our conceptions of the world (Halliday, 1994; Halliday and Matthiessen, 2004; Eggins, 2004). Thus, it can be cogently claimed that verbs are employed to denote all experience types known to human cognition, physical and concrete or mental and abstract.

D. Process Types and Transitivity

The basis of ideational metafunction of language is developed around phenomena that happen in the world experienced by human beings. Happenings do have causing factors or participants. They may have recipients. They inevitably happen within contexts, not in vacuum, thus they are/may be connected with the circumstances. Therefore, the basis of human experience, and subsequently the ideational potential of language are determined by what they perceive by phenomena and their related factors and participants. Lock (1996) defines *transitivity* as "the general term for the configuration of participants associated with different processes" (p. 73). Process in SFL is the semantic phenomenon realised by verbs.

Based on the type of the process, the accompanying participants are determined (Halliday and Matthiessen, 2004, inter alia). Halliday and Matthiessen (2004 and 1999) believe that there are six main process types in English. Figure 1 shows these six types schematically (Halliday and Matthiessen, 2004, p. 172).



Figure 1. The grammar of experience: process types in English.

The processes of the outer experience in the world are referred to as material processes, while those denoting the inner experience, or the experience of consciousness, are referred to as mental processes. The processes expressing the

link or relation between two fragments of experience are called relational processes. Verbal processes refer to those that reflect or are induced by inner experience and involve any type of saying or doing something with language. Behavioural processes refer to the outer manifestation of inner feelings, experience, intension or decisions, such as laughing. Finally, the existential process type refers to verbs by which phenomena of all kinds are simply recognised to 'be', or 'exist'. The point needs to be clarified that these are the main process types of most languages. There may exist minor types in some languages, and there is a degree of variance among languages in terms of the minor ones. Moreover, these notions are not crisp Aristotelian categories; rather, they are fuzzy notions and there may be some overlap among some type samples (Halliday and Matthiessen, 2004, pp. 170-174). Figure 2 presents the main elements of the system network of transitivity in English (Halliday and Matthiessen. 2004, p. 173).



Figure 2. Transitivity represented as system network.

E. Grammatical Metaphor in Processes

Transcategorisation is where the case of grammatical metaphor of processes, and therefore verbs, belongs in. There is, however, not a single route for all verbs to undergo grammatical metaphor. Compare the following examples (Halliday and Matthiessen, 1999, p. 242):

analyse verb: (analyst n./analysis n./analytic adj./analytically adv.)

behave verb: (behaviour n./ behavioural adj./ behaviourally adv./ behaviourist n.)

In some cases, fewer though, verbs are formed on the basis of other classes of words rather than being the platform for other cases of grammatical metaphor. The following are just a few instances (Halliday and Matthiessen, 1999, p. 242):

flake noun: (*flaky* adj./*flakiness* n./*flake* v.)

awake adj.: (awaken v./awakening n.)

Any judgment as to the originality (and therefore congruency) of either class of etymon should, for certain, be passed on extreme discretion and reliable diachronic (and semantic) evidence. This, however, may not be of prime importance on the sheer issue of transcategorisation. In terms of congruency or metaphoricity, on the other hand, (and (un) markedness as well) this plays an important part in determining different classes as more congruent (unmarked) or more metaphoric (marked). (In some cases, there would be no satisfactorily reliable level of certainty as of the congruency of either class of the same etymon.) In the case of grammatical metaphor through the vehicle of nominalisation, the case of verbs as the more congruent grammatical category accounts for the majority of ideational metaphors hinging around verbs based on the logic that the verbs at issue here are semantic denotations of processes. In the case of the noun-based verbs, however, the change applied on the base noun to transform it to a verb denotes an action happening to the concept and yet a nominalised grammatical metaphor is by no means out of the question.

III. THE PRESENT STUDY

Banks (2005) attributes the origin of grammatical metaphor in the form of nominalisation to Newton's Principia Mathematica from the seventeenth century, which was translated from the original Latin version into English in 1729 by Motte. These cases of nominalised grammatical metaphor run parallel to the Latin version. Banks (2005) believes that the influence of Latin on English in terms of nominalisation, while being rarely, if ever, considered, is worth careful study. This fact born in the mind, the conducted research appears disproportionately lagging behind. Some projects provide possible cases that grammatical metaphor can take (such as nominalisation) or common different functions they may serve in the sentences (e.g. Koptjevskaja-Tamm, 1993; Halliday, 1998). There are, however, some cases tackling

the phenomenon of grammatical metaphor in some details. Taverniers (2006), Sušinskienė (2004, 2008 and 2010), Farahani and Hadidi (2008), Sayfouri (2010) and Baratta (2010) are to be named as examples. Sušinskienė (2004, 2008 and 2010) considered the use of nominalised grammatical metaphor in scientific, political and media discourse, respectively. Taverniers (2006) sheds light on the semantic variation associated with grammatical metaphor. Farahani and Hadidi (2008) conducted a comparative study of the use of nominalisations between modern prose fiction and scientific discourse. Sayfouri (2010) compared the medical articles written and published by native and non-native authors and published in English and American journals on the one hand and in those published in Iran in terms of Swales'(1990) moves analysis as well as grammatical metaphor. Finally, Baratta (2010) considered the developmental path of nominalisation in the writings of undergraduate students. There, therefore, still exists the potential for more profound study of this linguistics phenomenon, particularly in the realm of genre analysis.

Based on the different natures of written and spoken language and the need for further research, the present research was inspired to compare the use of nominalisations in terms of their origin, frequency and the slots of the clause they serve to fill in in the discourse of science across the two written and spoken registers. Halliday (1989, 2007), Tillmann (1997), Aijmer and Stenström (2004), Wikberg (2004), Nelson, Balass and Perfetti (2005), Biber (2006), Miller (2006), McCarthy and Slade (2007) and Wichmann (2007) emphasise the different nature and aspects of written and spoken discourse, implicationally instilling the worthiness of further research on the differences between written and spoken discourse. The transient nature of spoken discourse versus the stability and permanence of writing, the time available (if any) to construct, revise and reconstruct the structures across the two language modalities, the nature and feature of interaction between the producing and the receiving ends in speech and writing, the level of formality and the different between written and spoken discourse in terms of the grammatical structures they employ frequently are some salient features of difference worth noting, which make research on discrepancies between them worth the while.

On the other hand, the research in most linguistic studies in general, and in grammatical metaphor in particular, has focused mainly on written discourse, leaving the spoken edition of language and the aspects it differs from the written edition almost totally unattended to. To be more precise, the studies on grammar, linguistic structures (mainly in structuralism), and genre analysis have frequently been based and conducted on written language marginalising the spoken corpora, whereas spoken data accounts for the bulk of the research on register analysis (Ferguson, 1996). Research projects spanning across the two register variations are, therefore, a rarity in the literature.

The present study, thus, has its inspiring origins in the remarkable potential of the research on the grammatical metaphor on the one hand and the different nature of spoken discourse from that of the written one and some scholars' admission that spoken language is, at least, as valuable as the written version on the other hand.

A. Material

Studer (2008) supports the single-genre data for the purpose of corpus analysis. He argues that if the corpus samples do not take adequate account of the properties of the genre(s) they are taken to be typical or representative of, the corpus is no longer possible to be used as a research tool. The point not to be neglected here is the grounds on which the corpus is to be developed, analysed and the findings to be interpreted. The corpus produced for a particular purpose would definitely be associated with problems or totally useless if used for a different goal. His suggestion for the use of single-genre corpora would not be of much value when the researcher addresses (a) general, not genre-specific, feature(s) of the language. But in most cases of research in the field of ESP that are focused on genres and their affinities and discrepancies, such a suggestion would be not only of great value but also indispensible.

The spoken scientific text type samples analysed herein were adopted from BBC radio programs in the form of downloadable audio files. They were downloaded from www.bbc.co.uk/podcasts over a period of about four years, from 2008 to 2011. It should also be noted that the programmes downloaded and analysed in this project were not all consecutive editions for practical download constraints and randomisation process.

The programmes chosen here were limited in terms of the number. This, per se, would seem an essential flaw for the samples on grounds of including few people's use of language, and therefore would be censured on account of limited generalisability, were it not for the type of the programmes included; they involved many different interviewees and consequently introduce the element of variety, and thus the basis for better reliability. (All programs normally involve at least two guests as interviewees.) Moreover, the diversity of the topics included in different editions of the programmes within each type of the text types extends the range of the areas to which the findings can be generalisable.

B. Transcription (raison d'être)

In order for the spoken language to lend itself to detailed analysis, it needs to undergo transcription to be transferred into the more stable format of writing. McEnery and Wilson (2001) consider the analysis of raw spoken data difficult and transcription a necessity for a successful analysis. Cohen, Manion and Morrison (2005) state that transcription, like many other sources of data, is a capable source in that it is the direct performance of the speaker translated from speech into writing. They consider it a useful way of avoiding being swamped by the abundance of the materials created. It gives the researcher the opportunity to review, re-read, categorise, summarise, interpret and analyse the data in the comfort and copiousness of his time and concentration. However, phonological and prosodic details would be left out in simple transcription. It should not be left short of heed that such compensatory techniques would be required for

prosodic studies and conversation analysis purposes. For the purpose of mere syntactic or lexical aspects of language analysis, such techniques would not be required or of much use.

C. Transcription Method

Halliday (1989) states that the proper way to transcribe a piece of spoken language is to include the possible pauses and thoughts that occur over the process of speech, similar to representing all the stages that a student goes through to prepare a composition. He names several methods for transcribing spoken language and points out their usefulness for research and pedagogical purposes. Notwithstanding the detailed complexity of the proposed methods, he notes that such extreme level of detailed transcription is not necessary for most purposes and that transcription into ordinary orthography is sophisticated enough for most research and pedagogical goals.

As mentioned above, there are many issues surrounding the issue of the quality and degree of details included. Regarding the present research, however, and the type of the analysis required here, the need for the detailed annotations is obviated and the verbatim lexical degree of transcription was adopted, and the annotations and phonetics symbols as well, are dispensed with.

It should be noted that grammatical dysfluencies were transcribed in fidelity to the original speech. These dysfluencies constituted some common types. In some cases, the speakers would, presumably unintentionally, repeat a lexical item while their mind is free to prepare and decide on the next items and the organisation and array they would appear in. Grammatical deviations from the standard grammar are also frequent occurrences. Such deviations would be either due to local dialects or the shortage of the mental processing capacity which seems to be excessively engaged by focus on the content or other issues. A third reason may presumably be the deficiencies with the speakers' language (often existing in non-native cases). The last case was quite rare in the data included in the present research. Moreover, such grammatical deviations are basically in the minor local scope and (almost never) cause any delay or breakdown in the communication. In the analysis phase, the grammatically correct versions were considered, and the repetitions, which spoken language (English in particular) abounds with, were, as well, considered single-occurrence, as expected in fully grammatical and standard version of the language. In case of the indistinct words or expression, the whole clauses were discarded from the analysis so that the analysis would be on complete clauses.

D. Spoken Scientific Discourse

The samples of spoken language with scientific content were chosen from BBC's science programmes. The source programmes were *Science in Action, Material World, Digital Planet, Medical Matters, Health Check* and *In Our Time*, downloaded from <u>www.bbc.co.uk/podcasts</u> over a period of three years, from 2008 to 2010. The first two programme deal with the most recent scientific breakthroughs in different areas of science and their applications. The third one sheds light on the recent issues and advancements in the field of computer-based technology. The fourth and fifth are normally related to the advancement in medical sciences and health issues and problems in real life. These programmes and their approximate length are listed below:

Science in Action: 4000 - 4500 words each, 16 programmes, 69,480 words

Medical matters: 5000 - 5500 words each, 16 programmes, 85,772 words

Material World: 5000 - 5500 words each, 16 programmes, 86,897 words

Health Check: 4500 - 5500 words each, 16 programmes, 72,354 words

Digital Planet: 4500 - 5500 words each, 16 programme, 81,632 words

The total number of the words of these programmes amounted to 396135. In order to round the number off at approximately 400,000 words, another programme, which was an edition of *Science in Action* (4384 words), was randomly chosen and then added to the list given above. The total number of the words included in the category of spoken scientific discourse, therefore, was 400,519. Some sentences were discarded at random to bring the number even closer to the round number.

E. Written Scientific Discourse

The process of the selection in this category was identical to the previous one (based on page numbers, with random deletion of some clauses). The only difference was the sources which the selection was made from. The following the list of the sources from which the written scientific samples were selected:

1. Scientific American, Online exclusive issue, no. 12, Extreme Physics, 2004.

2. Scientific American, 2009 (all issues included).

3. Life: The Science of Biology (ninth edition), 2011, by Sadava, D., Hillis, D. M., Heller, H., C., and Berenbaum, M. R., published by Sinauer Associate, Inc. and W. H. Freeman and Company, Sunderland and Gordonsville (respectively): USA.

4. The Theories of Chemistry, 2003, by Boeyens, J. C.A., published by Elsevier, Amsterdam.

5. Basics and Highlights in Fundamental Physics (Proceedings of the International School of Subnuclear Physics), 2001, edited by Zichichi, A., published by World Scientific: Singapore.

6. Fundamentals of Physics, 9th edition, 2011, by D. Halliday and R. Resnick and J. Walker, published by John Wiley and Son, Inc.: USA.

F. Some Notes on the Method

In the case of inversion wherein nominalisations were involved, the judgment was based on the uninverted version. As an instance, the nominalised 'awakening' was considered to be in post verbal position in the extract: 'a spring awakening perhaps we can do with some kind of spring after all'. The logic behind the choice was the semantic basis which subordinates an object to a verb, not vice versa. In other words, the present study hinges around the framework of transitivity rather than thematic (theme-rheme) structure. However, in detailed semantic considerations, the original array of words would be of higher priority.

Appositives were considered in the category of the word(s) they were based on. In 'Speed, for example, is defined in terms of the base quantities, <u>length</u> and time, and their base standards', '<u>length</u>' would be considered an object of preposition. The substitution of verbs were likewise considered as the words they substituted for on grounds that the substitutions are in fact forms denoting specific semantic concepts belonging to the words they represent. In case of the verbs, the common substituting element is 'do' and its inflections. A similar occurrence, mainly in spoken language, is ellipsis, or substitution by zero elements. Ellipses are normally based on the physical, or less commonly, linguistic context. 'Don't!', for instance, relies on the physical context for its sematic interpretation. Some other cases of ellipses are readily possible to be interpreted through simple reliance of prevalence or linguistic context. A speaker's utterances like 'Good!' or 'Fine' would easily be interpreted as the reduced forms of 'It's good!' or 'It's fine!'. This last type is, however, absent from the formal written register.

In case of the occurrence of formulae in written scientific passages, in chemistry and physics passages for instance, they were discarded and therefore not included in the word count. This was done on grounds of formulae being more of merely notational than linguistic nature, and the present research did not aim at multisemiosis or intersemiosis.

The repetitions of linguistic elements or structures were treated as single occurrence. The case of repetition in spoken language, in the samples analysed here, appeared to be of two types; they occurred either for the purpose of providing their mind some time to prepare the following words in the stream of speech, which is typically limited to single item structures, or for the purpose of adding emphasis to the repeated structure, which may include a whole clause. In either case, the sheer occurrence is a phenomenon of a pragmatic rather than purely linguistic nature, and the repetitive cases were considered a single occurrence.

G. Point of Departure

The initial analysis of texts without a general sketch of what to expect and how to categorise the cases of deverbal nominalisations and verbs is, in nature, utterly qualitative, and, therefore, demands extensive amount of time and unaffordable level of mental capacity to avoid the definite confusion created by such large amount of qualitative data at hand, which is beyond human mental capacity. A possible solution would be resorting to software realm. Software programs devised for the purpose of text analysis are yet at preliminary stages, with the complexity and abstraction level of language use and meaning in view. Therefore, analyses beyond the clear-cut classes of words and clauses either require human intervention or totally manual analysis of language samples. The limited human mental capacity, however, necessitates the definition of a clear-cut and manoeuvrable framework as a basis for (further) effective analysis.

To the end already explained, first, 10,000 word samples were analysed for the purpose of primary analysis. The word categories at issue, namely deverbal nominalisations and verbs, were considered in terms of their types and the structures wherein they were used. Once the instances of the word types at issue in the preliminary samples were considered, the resultant types were classified and tabulated, to a degree of comprehensiveness as to encompass all probable instances of a specific word sample in distinct groups and to an extent of detail as not to perplex the process of analysis and handling the groups and the data thus obtained. At next stage, the rest of the text samples intended for analysis were analysed in accord with the already based classifications and the frequency of each cell in the table was added, in an accumulative manner, to the results from the preliminary stage of analysis. The functional systemic typology of processes (Halliday and Matthiessen, 2004, for instance) was used for the classification of verb types and types of nominalisations, but as to the slots the nominalisations fill, four main categories emerged from the preliminary study: *pre-verbal, post-verbal, object of preposition* and finally as *modifier*.

In the case of verbs, being finite or non-finite was another factor of analysis. Finite clauses are those housing a tense of verb, like all independent and dependent clauses, and the non-finite clauses are those lacking a clear tense element. Participle clauses, either present or past, and infinitive clauses are all non-finite. This point was considered based on the grounds that non-finite clauses, especially participle clauses, are thought of as adding syntactic complexity to the discourse, and therefore a feature mainly prevalent in written academic register.

As it is common for similar studies in the field of discourse analysis, the data in this research did not undergo any statistical analysis; the figures are presented in the form of raw frequencies. In cases the number of the words analysed in a specific text type exceeds a million, the frequencies would be standardised as the number of a specific type of words in a million words. Here, therefore, since the number of the words in each type of text is 400,000, the results are reported in raw frequencies.

IV. RESULTS

The cumulative results of the analysis of the two discourse types are presented in the following tables. TABLE 1 presents the frequency of deverbal nominalisations in spoken scientific discourse and TABLE 2 shows their frequency in written scientific discourse. In these two tables, the frequency of nominalisations in terms of the underlying process types and the four slots they occupy in the sentence is provided for spoken and written scientific discourse types.

DEVERBAL NOMINALISATIONS IN SPOKEN SCIENTIFIC DISCOURSE.					
	Pre-verbal	Post-verbal	Obj. of prep.	Modifier	N
Material	623	1052	4479	187	6341
Behavioural	52	391	104	27	574
Mental	572	1049	176	85	1882
Verbal	233	182	664	39	1118
Relational	43	138	214	28	423
Existential	8	13	17	0	38
Ν	1531	2825	5654	366	10376

1	AB.	LE.	1

TABLE 2.			
DEVERBAL NOMINALISATIONS IN WRITTEN SCIENTIFIC DISCOURSE.			

	Pre-verbal	Post-verbal	Obj. of prep.	Modifier	N
Material	3782	4081	4763	576	13202
Behavioural	98	134	182	2	416
Mental	566	1045	1118	4	2733
Verbal	231	747	715	87	1780
Relational	10	26	29	5	70
Existential	8	19	34	3	64
Ν	4695	6052	6841	677	18265

The frequency of verbs in terms of their process types in each of the two discourse types are provided in TABLES 3 and 4 for the purpose of comparison. Their frequency in the two clause types of finite and non-finite is also part of the data provided in these tables.

TABLE 3.
PROCESS TYPES IN SPOKEN SCIENTIFIC DISCOURSE.

	Finite	Non-finite	Ν
Material	15307	4917	20224
Behavioural	2165	874	3039
Mental	7793	578	8371
Verbal	3539	1335	4874
Relational	16203	1402	17605
Existential	1529	4	1533
Ν	46536	9110	55646

TABLE 4.	
----------	--

PROCESS TYPES IN WRITTEN SCIENTIFIC DISCOURSE

	Finite	Non-finite	N
Material	9326	4376	13702
Behavioural	76	35	111
Mental	2737	864	3601
Verbal	2683	1182	3865
Relational	7058	1932	8990
Existential	503	78	581
Ν	22383	8467	30850

As it can be seen in the tables, there are considerable differences among various process types (as verbs or nominalisations) and their accommodating structures across the two modalities (spoken and written). Probably the first striking point would be the difference in the frequency of verbs. Spoken scientific samples included 55646 verbs in all, while in the written counterpart, the number is 30850. Another noticeable point across the two modalities is the difference between finite and non-finite use of verbs; in spoken discourse only 16.4% of verbs occur in non-finite clause, while in written discourse, non-finite clauses account for 27.5% of all verbs used. In terms of nominalisation frequency, written register stands in a higher position than the spoken register does (18265(written) vs. 10376(spoken)), while in terms of verb frequency, the converse applies (55646 (spoken) vs. 30850 (written)).

The frequency of deverbal nominalisations appears to vary process type and the slots they fill in the sentence. The common overall ranking of frequency in both registers is *object of preposition, post-verbal, preverbal* and *modifier*. However, the values were not of the same proportion across the four slots for the two text types; in written discourse, the first three slots enjoyed values closer to one another than in spoken discourse. In other words, the difference between the two text types in terms of nominalisation frequency increases from the first slot down to the last. It is worth mentioning that in spoken scientific discourse, this trend applies to the total number of deverbal nominalisations, and to

material, relational and existential process types; behavioural, mental and verbal process type nominalisations deviate from this overall pattern. In written scientific discourse, the overall pattern applies to all process types' nominalisations.

The differences between the two registers appear to vary considerably across process types. All in all, material process type accounts for a considerable part of the processes used in the two registers, both in verbs and nominalisations. The interesting point is that relational processes rank second in terms of the verbs used, in both written and spoken discourse type, but are relegated to the fifth place just above that of existential process type (in both written and spoken text types).

V. CONCLUSION AND DISCUSSION

Spoken language turns out to have more verbs in a given scope of a sample compared to the written language. With the scope of the samples in the two modalities being the same, this fact shows that spoken language accommodates more clauses and propositions. It follows that there are fewer words in each sentence and clause in spoken language. The transient and online nature of speech requires that the speaker use shorter clauses in order not to inundate the listeners' minds with too much load to be decoded at the same pace as the speakers'. Therefore, as logically expected, speech includes clauses easier to encode and decode. With written language, the language user has more time to reflect upon different parts of clauses and passages and usually makes them longer, more refined, sophisticated, and semantically and syntactically loaded.

The higher frequency of non-finite clauses in written discourse compared with spoken discourse would logically indicate a higher level of abstraction and complexity in written language. Participle clauses and infinitive clauses add to complexity level of the language in that they make a proposition syntactically a subordinate of another clause. Moreover, due to their complexity and the demand they place on the mind of both participants in a communication (speaker or writer on the one hand and listener or reader on the other), participle clauses are generally considered mainly a feature common to written formal register rather than spoken. In the samples considered herein, however, there is even a higher likelihood of participle clauses being used than in informal spoken register. Formality can safely be considered a factor bringing spoken language closer to written language, while colloquial, informal language or slang are further away from the written edition on the formality continuum. In the case of nominalisations, the same underlying issue applies; the frequency of nominalisations in spoken language is lower than that in written language, and so is the semantic complexity and lexical density created by them.

Material process type constitutes more than 60% of nominalisations in spoken and written discourse type investigated here (61% and 72%, respectively) while the frequency drops to 36% and 44% in the case of verbs used in spoken and written text samples, respectively. An opposite trend is salient in the case of relational processes. They rise from almost a negligible minority in the realm of nominalisation to the conspicuous second most frequent type among verbs in both modalities (31.6% and 29% of verbs in spoken and written samples, respectively). A logical corollary of these changes in frequency is that nominalisations fall mainly in material process type leaving relation process type out by a wide margin. Language reflects human experiences and conceptions. It is safe to assume that most of the events and phenomena human beings experience are of material and tangible nature (doing or happening). The rest of the process types can be considered to have mainly evolved to handle the concepts denoting material experience; the relational process type denotes how human experience can be related with one another and handled. Mental and behavioural types are used for inner experiences, conceptions and feeling which have their origin within human beings rather than outside, and as the results show, are limited in scope compared with material type.

The precedence of nominalisation occurrence in the post-verbal position over that in pre-verbal position could provide good clues as to the nature of language processing in the mind of human beings. Nominalisations, as the condensed form of clauses, add to the semantic load and lexical density of language. It can, therefore, be inferred that human mind tends to implement and maintain verbal communication through simple points of departure for clauses and to expand and elaborate on them to provide more pertinent information. Since the information structure of given and new overlaps in absolute majority of cases the theme-rheme structure, it can safely be assumed that human being naturally tend to start their clauses and sentence with smaller pieces of information and then proceed to provide more (elaborate) information following the given information as a prompt. In the case of written language, on the other hand, the mind is set free from the constraints inherent in speech and is therefore provided with more freedom to manoeuvre on the point of departure (the difference between pre-verbal and post-verbal nominalisations is less in written language than that in spoken discourse in terms of their proportion).

There are some caveats to be considered in all studies of language analysis in all forms including genre analysis, register analysis, ESP speculations and even the division of language into spoken and written modalities. Bhatia (2001) and Hyland (2002) note that any classification of language into different genres, registers or ESP taxonomies are subject to a level of continuity and overlap, and therefore involve some degree of variation and dynamicity. This is due to the fact that language reflects the reality and reality does not necessarily include discrete concepts and sharp borders among its entities, and consequently is inherently full of fuzziness and indeterminacies. As a result, the system reflecting the reality would inevitably involve many cases of fuzziness and indeterminacy. Bennet (2009) and Martin (1993) also point out the fact that the distinction presumed to exist among academic discourse in terms of humanities or sciences actually fade in reality into subtle and gradual continuity. Hammond (1990) has a similar viewpoint towards the sharp

dichotic distinction between spoken and written language and assumes an inherent relationship between them. The present study, therefore, is no exception and does not have any definitive claims as to its findings and, by the same token, considers its findings and implications not sharp absolute distinctions but rather treats them as descriptions of language samples lodged on different spots of a continuum determined by a grid of a myriad of factors ranging from content, the relationship between the interlocutors involved, the modality, the formality level, users' idiosyncrasies... existent in the real world and the language employed to describe it.

REFERENCES

- [1] Aijmer, K. and Stenström, A. B. (2004). Discourse patterns in spoken and written corpora. In K. Aijmer and A. B. Stenström (Eds.) *Discourse Patterns in Spoken and Written Corpora*. Amsterdam: John Benjamins Publishing Company.
- [2] Baake, K. (2003). Metaphor and Knowledge: The Challenges of Writing Science. New York: State University of New York Press.
- [3] Banks, D. (2005). On the historical origins of nominalized process in scientific text. *English for Specific Purposes*. 24, pp. 347-357. DOI: 10.1016/j.esp.2004.08.002.
- Baratta, A. M. (2010). Nominalization development across an undergraduate academic degree program. *Journal of Pragmatics*, 42, pp. 1017-1036. DOI: 10.1016/j.pragma.2009.08.007.
- Bennet, K. (2009). English academic style manuals: A survey. *Journal of English for Academic Purposes*. 8, pp. 43-54. DOI: 10.1016/j.jeap.2008.12.003
- [6] Bhatia, V. K. (2001). Analysing genre: some conceptual issues. In M. Hewings, (Ed.) *Academic Writing in Context*. Birmingham: Birmingham University Press.
- [7] Biber, D. (2006). University Language: A corpus-based study of spoken and written registers. Amsterdam: John Benjamins.
- [8] Eggins, S. (2004). An Introduction to Systemic Functional Linguistics (2nd Ed.). London/New York: Continuum.
- [9] Fairclough, N. (2003). Analysing Discourse: textual analysis for social research. London/New York: Routledge.
- [10] Farahani, A. A. and Hadidi, Y. (2008). Semogenesis under Scrutiny: Grammateical Metaphor in Science and Modern Prose Fiction. *Iranian Journal of Applied Linguistics (IJAL)*. 11(2) pp. 51-82.
- [11] Halliday, M. A. K. (1985). An Introduction to Functional Grammar. London: Edward Arnold.
- [12] Halliday, M. A. K. (1989). Spoken and Written Language. Oxford: Oxford University Press.
- [13] Halliday, M. A. K. (1994). An Introduction to Functional Grammar (2nd ed.). London: Edward Arnold.
- [14] Halliday, M. A. K. (1998). Things and relations: Regrammaticising experience as technical knowledge. In J. R. Martin and R. Veel (Eds.), *Reading Science: Critical and functional perspectives on discourses of science*. London/New York: Routledge.
- [15] Halliday, M. A. K. (2003). Introduction: On the "architecture" of human language. In J. J. Webster (Ed.) On Language and Linguistics: Volume 3 in the Collected Works of M. A. K. Halliday. London/New York: Continuum.
- [16] Halliday, M. A. K. (2007). Differences between Spoken and Written Language: Some implications for Literacy Teaching. In J. J. Webster (Ed.) Language and Education: Volume 9 in the Collected Works of M. A. K. Halliday. London/New York: Continuum.
- [17] Halliday, M. A. K. and Hasan, R. (1976). Cohesion in English. London/New York: Longman.
- [18] Halliday, M. A. K. and James, Z. J. (1993). A Quantitative Study of Polarity and Primary Tense in the English Finite Clause. In J. M. Sinclair, M. Hoey and G. Fox, *Techniques of Description: Spoken and written discourse*. London/New York: Routledge.
- [19] Halliday, M. A. K. and Matthiessen, C. M. I. M. (1999). Construing Experience through Meaning: A Language-based Approach to Cognition. London/New York: Continuum.
- [20] Halliday, M. A. K. and Matthiessen, C. M. I. M. (2004). An Introduction to Functional Grammar. (3rd ed.). London: Edward Arnold.
- [21] Hammond, J. (1990). Oral and Written Language in Educational Context. In M.A.K. Halliday, J. Gibbons and H. Nicholas (Eds.), *Learning, Keeping and Using Language, Volume I.* Amsterdam/Philadelphia: John Benjamins.
- [22] Hyland, K. (2002). Specificity revisited: how far should we go now?. English for Specific Purposes. 21, pp. 385-395.
- [23] Koptjevskaja-Tamm, M. (1993). Nominalizations. London/New York: Routledge.
- [24] Martin, J. R. (1993). Technicality and Abstraction: Language for the Creation of Specialized Texts. In M.A.K. Halliday and J. R. Martin, (eds.). Writing Science: Literacy and Discursive Power. London/Washington DC: The Falmer Press.
- [25] Martin, J. R., Matthiessen, C. M. I. M. and Painter, C. (1997). Working with Functional Grammar. London: Edward Arnold.
- [26] Matthiessen, C. M. I. M., Teruya, K. and Lam, M. (2010). Key Terms in Systemic Functional Linguistics. London/New York: Continuum.
- [27] McCarthy, M. and Slade, D. (2007). Extending Our Understanding of Spoken Discourse. In J. Cummins and C. Davison (Eds.) International Handbook of English Language Teaching. New York: Springer.
- [28] McEnery, T. and Wilson, A. (2001). Corpus Linguistics. Edinburgh: Edinburgh University Press.
- [29] Miller, J. (2006). Spoken and Written English. In B. Arts and A. McMahon (Eds.) *The Handbook of English Linguistics*. Oxford: Blackwell Publishing.
- [30] Nelson, J. R., Balass, M., and Perfetti, C. A. (2005). Differences between written and spoken input in learning new words. Written Language & Literacy. 8(2) pp. 25–44.
- [31] Sayfouri, N. (2010). SFL and ESP Genre Analysis of English Research Articles in Iranian and English-American Medical Journals: A Contrastive Study. Unpublished PhD dissertation in TEFL. Department of English Language and Literature, The University of Tehran, Tehran, Iran.
- [32] Steiner, E. (2002-2003). Ideational Grammatical Metaphor: Exploring some implications for the overall model. *Languages in Contrast*. 4(1) pp. 137-164.
- [33] Studer, P. (2008). Historical Corpus Stylistics: Media, Technology and Change. London/New York: Continuum.
- [34] Sušinskienė, S. (2004). Grammatical Metaphor in Scientific Discourse. KALBOTYRA 54(4), pp. 76-83.

- [35] Sušinskienė, S. (2008). Grammatical metaphor as ideological micro-components of political discourse. *Filologija*. 13 pp. 129-138.
- [36] Sušinskienė, S. (2010). Nominalization as a cohesive device in British newspaper editorial. *Filologija*. 15 pp. 142-150.
- [37] Swales, J. M. (1990). Genre Analysis: English in Academic and Research Settings. Cambridge: Cambridge University Press.
- [38] Taverniers, M. (2006). Grammatical Metaphor and Lexical Metaphor: different perspectives on semantic variation. *Neophilologus*, 90 pp. 321-332.
- [39] Thompson, G. (1996). Introducing Functional Grammar. London: Edward Arnold.
- [40] Thompson, G. (2004). Introducing Functional Grammar (2nd ed.). London: Edward Arnold.
- [41] Tillmann, H. G. (1997). Eight Main Differences between Collections of Writing and Spoken Language Data. Forschungsberichte des Instituts f ür Phonetik und Sprachliche Kommunikation der Universit ät M ünchen (FIPKM) 35 pp. 139-143.
- [42] Wichmann, A. (2007). Corpora and spoken discourse. In R. Facchinetti (Ed.) Corpus Linguistics 25 Years on. Amsterdam: Rodopi.
- [43] Wikberg, K. (2004). English metaphors and their translation: The importance of context. In K. Aijmer and A. B. Stenström (Eds.) Discourse Patterns in Spoken and Written Corpora. Amsterdam: John Benjamins Publishing Company.



Mohammad Hussein Norouzi was born in Tabriz, Iran, in August 1979. He did his BA in English language and literature at the University of Tabriz, Tabriz, Iran, graduating in 2001. He finished his MA at the University of Tehran, Iran, in 2004, in the field of TEFL. At the moment, he is a PhD candidate at the University of Tehran in the same field as his MA's.

He has taught English conversation in the Open Centre for Cultural Educations, an affiliate of the University of Tehran from 2002 till 2005. He taught IELTS in an affiliate institute of The University of Tehran, Farhikhtegan-e Daneshgah from 2005 up till 2010. Currently, he is a Lecturer at the University of Tehran, Tehran, Iran. He has been in the last position since 2006. He has published an article titled 'Self-Assessment' in ILI Language Teaching Journal, Volume 4, No. 1, year 2008, and has co-authored an article

'Readability Formulas and Cohesive Markers in Reading Comprehension' in Theory and Practice in Language Studies, Volume 1, No. 8, pages 1005-1010, August 2011. Discourse Analysis constitutes his main area of interest in Applied Linguistics.

Ali Akbar Khomeijani Farahani is an assistant professor of English and Linguistics in the English department of the University of Tehran. He has been teaching linguistics and English in this department for more than 21 years. His research interests include discourse analysis and systemic functional linguistics. He has taught extensively in these areas at MA and PhD levels.

He received his PhD in Linguistics from Leeds University UK in 1990. He received his MA in the same field in 1986 from the same University. He received his BA in the English language and literature from the faculty of humanities at the University of Tehran in 1980. He did his military service from 1981 to 1983. Then he got a job in NIOC as a senior buyer. He held this job until 1985 when he went to UK to further his studies.

Dr. Khomeijani Farahani has already published quite a few articles on different topics related to linguistics and foreign language teaching and a book. Three of the most recent articles and the book are: Articles: Do Different Textual Enhancement Formats Have Differential Effects on the Intake of English Subjunctive Mood? (2012). Academy publisher Finland. Foreign Language Learners' processing of Relative Clause Ambiguity (2011). Iranian EFL Journal Tehran Iran. Cognitive Task Complexity and L2 Narrative Writing Performance (2011). Journal of Language Teaching and Research. Finland. Books: Essential Roots: Prefixes and Suffixes (Tehran, Iran: Jungle Publications, 2012)

Davood Borzabadi Farahani was born in Tehran, Iran in 1962. He received his university degrees from two state universities in the capital city of Tehran in Iran. More specifically, he received his B.A. degree in English translation from Allame Tabatabee University, his M.A. in TEFL from the University of Tehran and his Ph.D. from Allame Tabatabee University in the years 1989, 1991, and 2000, respectively.

He has been teaching English at the University of Tehran and other language institutes for the past twenty years. In fact, he is a faculty member of the English Department of the College of Foreign Languages and Literatures at the University of Tehran. He is also a test developer at a state-run testing organization named the National Organization for Educational Testing (NOET).

Dr. Borzabadi has so far published several research articles in Iranian applied linguistics journals. Among such articles are The Teacher Parameter in the Postmethod Era (2007), Evaluation of the Construct Validity and Reliability of Language Learning Styles Instruments (2008), and Investigation the Construct Validity and Reliability of Oxford's Strategy Inventory for Language Learning (SILL) (2009).