CROSSROADS

A Journal of English Studies

ISSUE 21
2/2018

An electronic journal published
by The University of Bialystok

CrossRoads

A Journal of English Studies

ISSUE 21

2/2018

An electronic journal published by The University of Bialystok

CROSSROADS
A Journal of English Studies

Publisher:
The University of Bialystok
The Faculty of Philology
Department of English
ul. Liniarskiego 3
15-420 Białystok, Poland
tel. 0048 85 7457516
crossroads@uwb.edu.pl
www.crossroads.uwb.edu.pl

e-ISSN 2300-6250
The electronic version of *Crossroads. A Journal of English Studies* is its primary (referential) version.

Contents

TRANSLATION STUDIES

KRZYSZTOF HEJWOWSKI
University of Warsaw

DOI: 10.15290/cr.2018.21.2.01

Applied Linguistics and Translation Studies

Abstract: The article deals with the relation between translation studies and linguistics, in particular applied linguistics. Some facts from the history of translation studies and applied linguistics are presented (James Holmes's famous paper delivered at the III International Congress of Applied Linguistics, the beginnings of the Institute of Applied Linguistics at Warsaw University). A few definitions of applied linguistics and translation studies are discussed. In conclusion the author's view on the status of translation studies is expounded.

Keywords: translation, applied linguistics, translation studies.

In 1972 at the Third International Congress of Applied Linguistics James Holmes delivered his famous paper entitled "The Name and Nature of Translation Studies", in which he established the name of the discipline and classified it as an empirical discipline, as such having two major objectives:

- to describe particular phenomena in the world of our experience,
- to establish general principles by means of which they can be explained and predicted.

Holmes also divided translation studies into "pure" and applied (Holmes 2000: 176). He did not, however, state that translation studies was a subdiscipline of applied linguistics. He delivered his lecture at a congress of applied linguistics because translation conferences at that time were very scarce (and translation departments or faculties – practically non-existent).

In the same year 1972 the Institute of Applied Linguistics was founded at the University of Warsaw. The Institute trains future teachers, translators and interpreters. Each student follows courses in two foreign languages (at present out of 7: English, German, French, Spanish, Russian, Swedish and Japanese) and two specializations: foreign language teaching and translation. Thus, the name of the Institute and its curriculum might suggest an idea that applied linguistics contains those two main branches: language teaching and translation.

What is really the relationship between applied linguistics and translation studies? Let us first look at some dictionary definitions. In the "Introduction" to *Longman Dictionary of Language Teaching and Applied Linguistics* (Richards et al. 1992) its authors inform that

> This dictionary includes the core vocabulary of both language teaching and applied linguistics. [...]
> For the purposes of this book, "applied linguistics" refers to the practical applications of linguistics and language theory, and includes terms from the following areas of study:
>
> - introductory linguistics, including phonology, phonetics, syntax, semantics and morphology
> - discourse analysis
> - sociolinguistics, including the sociology of language and communicative competence
> - psycholinguistics, including first and second language acquisition, contrastive analysis, error analysis and learning theories (Richards et al. 1992: vii)

It would seem that for the authors of the dictionary language teaching and applied linguistics are two separate disciplines (Translation is not mentioned). On the other hand, the entry "applied linguistics" contradicts the above statement:

> applied linguistics
>
> 1. the study of second and foreign language learning and teaching. [sic!]
> 2. the study of language and linguistics in relation to practical problems, such as lexicography, translation, speech pathology, etc. Applied linguistics uses information from sociology, psychology, anthropology, and information theory as well as from linguistics in order to develop its own theoretical models of language and language use, and then uses this information and theory in practical areas such as syllabus design, speech therapy, language planning, stylistics, etc. (Richards et al. 1992: 19)

In the second subentry translation is mentioned as a "practical problem" – presumably dealt with by "the study of language and linguistics".

The dictionary contains entries **translation** ("the process of changing speech or writing from one language [...] into another...") and **translation equivalence** ("the degree to which linguistic units [...] can be translated into another language without loss of meaning") (Richards et al. 1992: 389) but not "translation studies". The definitions of translation and translation equivalence prove how little the authors have to say about translation. Even though the dictionary was first published in 1985, the ignorance of translation theory is appalling.

A corresponding Polish dictionary bears the title *Podręczny słownik językoznawstwa stosowanego* ('Desk dictionary of applied linguistics') and a subtitle "Dydaktyka języków obcych" ('Foreign language teaching') (Szulc 1984). In the introduction the author states that

> Termin „językoznawstwo stosowane" rozumieć należy w tym kontekście w jego tradycyjnym, zaznaczonym w podtytule, ujęciu, tzn. jako zastosowanie badań językoznawczych do dydaktyki języków obcych. ("The term 'applied linguistics' should be understood in this context in its traditional, stressed in the subtitle, meaning, i.e. as applying linguistic research to foreign language teaching") (Szulc 1984: 5)

Consequently the dictionary contains no entries connected with translation. Only "machine translation" is mentioned in the entry **językoznawstwo stosowane** (applied linguistics):

> Dział językoznawstwa zajmujący się możliwościami praktycznego zastosowania (np. w dydaktyce języków obcych lub informatyce) osiągnięć takich dyscyplin jak językoznawstwo, socjologia, socjolingwistyka, psychologia, psycholingwistyka, antropologia i in. [...]
> W kręgu zainteresowań językoznawstwa stosowanego znajduje się również zagadnienie afazji, tłumaczenia maszynowego oraz telekomunikacji.
> ("A branch of linguistics dealing with practical application (e.g. in foreign language teaching or computer technology/information science) of the findings of such disciplines as linguistics, sociology, sociolinguistics, psychology, psycholinguistics, anthropology and others. [...] The discipline is also interested in such problems as aphasia, machine translation and telecommunication.") (Szulc 1984: 104-105)

The definition seems to take it for granted that sociology, psychology or anthropology have no applied branches and therefore their findings have to be utilized by applied linguistics. Thus applied linguistics becomes some meta-discipline though its "fields of application" are rather narrow.

W. Grabe in an introductory article to *The Oxford Handbook of Applied Linguistics* (Kaplan 2002) defines applied linguistics as "a practice-driven discipline that addresses language-based problems in real-world contexts" (Grabe 2002: 10). Translation is mentioned as a "sub-field of study":

> Applied linguistics generally incorporates or includes several further identifiable sub-fields of study: second language acquisition, forensic linguistics, language testing, corpus linguistics, lexicography and dictionary making, language translation, and second language writing research. (Grabe 2002: 11)

Two things deserve our attention: first, placing translation somewhere between language testing and second language writing, and the very phrase "language translation", which reveals lack of knowledge about the issue. One does not translate languages, one translates texts. The marginal position occupied by translation in such a vision of applied linguistics is well reflected by the size of contributions devoted to translation in the whole volume: there are two articles (one entitled

"Translation", the other – "Interpretation") occupying some 30 pages out of 630. Other sections of the book include: "The four skills: speaking, listening, reading, and writing", "Discourse analysis", "The study of second language learning", "The study of second language teaching", "Variation in language use and language performance", "Bilingualism and the individual learner", "Multilingualism in society", "Language policy and planning", "Language assessment and program evaluation", "Technological applications in applied linguistics". The last chapter contains two articles: "Directions in automated essay analysis" and "Computer-assisted language learning". No mention of CAT tools or machine translation. All this very clearly shows that applied linguists either exclude translation from their field of interest or treat it as a far periphery of their discipline and usually do not say anything interesting about it.

Let us now look at the definitions provided by the most popular source of knowledge, Wikipedia:

> Applied linguistics is an interdisciplinary field of linguistics which identifies, investigates, and offers solutions to language-related real-life problems. Some of the academic fields related to applied linguistics are education, psychology, communication research, anthropology, and sociology.
> Major branches of applied linguistics include bilingualism and multilingualism, conversation analysis, contrastive linguistics, sign linguistics, language assessment, literacies, discourse analysis, language pedagogy, second language acquisition, language planning and policy, interlinguistics, stylistics, pragmatics, forensic linguistics and translation. (Wikipedia "applied linguistics")

The second part of the definition seems to confuse disciplines with objects of study – different branches of linguistics, discourse and conversation analysis, pedagogy, stylistics and pragmatics on one hand, and literacies, second language acquisition, language planning and policy, translation on the other. It may be instructive to compare the definition with that of "translation studies":

> Translation studies is an academic interdiscipline dealing with the systematic study of the theory, description and application of translation, interpreting, and localization. As an interdiscipline, Translation Studies borrows much from the various fields of study that support translation. These include comparative literature, computer science, history, linguistics, philology, philosophy, semiotics, and terminology. (Wikipedia "translation studies")

Curiously, the definition does not mention psychology, sociology or anthropology among the "fields of study that support translation". It is also interesting to notice how both definitions formulate their genus proximum: "an interdisciplinary field of linguistics" and "an academic interdiscipline".

I can understand the word "interdisciplinary" – "of or relating to more than one branch of knowledge" (*The New Oxford Dictionary of English*), where "of" probably means "belonging to". I am not sure whether I understand the word "interdiscipline". Traditional dictionaries do not

contain this entry. Is it a discipline after all? An "interdisciplinary discipline"? A definition from Wikipedia does not help much:

> The term interdiscipline [...] means an organizational unit that involves two or more academic disciplines, but which have [?] the formal criteria of disciplines such as dedicated research journals, conferences and university departments. It is related to interdisciplinarity, but it is a noun used for a certain kind of unit (academic discipline). As shown in the example of demography below a field may be both a discipline and an interdiscipline at the same time. The example of information science demonstrates that a field may be regarded as a discipline in some countries but an interdiscipline in other countries. (Wikipedia "interdiscipline")

This pseudo-definition does not answer our question. In fact, it does not answer any questions. We are still left with our query concerning the status of translation studies and its relationship with applied linguistics.

What are the criteria of acknowledging that a field of study is an independent academic discipline? First of all it should have a serious, important, complex and complicated object of study. Is translation such an object? It is definitely serious and important. We are surrounded by translations – they are everywhere. We read translated books and documents, web pages and user's manuals, we watch translated films and TV programmes, play translated games, listen to translated news from other countries. Most of our knowledge of the world comes from translations. So, after a moment of thought, we have to admit that translations are important and omnipresent. But is translation complex and complicated? Recently I have kept repeating I. A. Richards' words: "Translation may very probably be the most complex event yet produced in the evolution of the cosmos" (Richards 1953: 250). I keep repeating them because they are very true and because hardly anybody realizes the complicated nature of translation. Most people believe that translating is about replacing words from one language with their dictionary equivalents from the other language (cf. the "definition" of translation in *Longman Dictionary of Applied Linguistics* quoted above). Translation is not replacing words with other words. Translation consists in reproducing the mental structures signalled by text A in language a, and then producing text B in language b that will make it possible for users of language b to reproduce as much of those mental structures as possible. Merely to understand what translation really is one needs to practise (preferably to practise and study) it for a few years. In order to become a professional translator more years of practice (preferably of practice and study) are needed.

Thus, we have a serious, important, complex and complicated object of study. We have thousands of scholars dealing with it. We have academic departments, numerous conferences, journals, plenty of students eager to study translation. In my opinion it is enough to recognize translation studies as an independent academic discipline. Does the word independent mean that it has no contacts with other disciplines, that it does not owe anything to other fields of study? Of course not. There are no such disciplines in contemporary "interdisciplinary" world. Translation studies

definitely draws on psychology, sociology, anthropology, literary studies, philosophy... And obviously on linguistics. This "dependence" on other fields of study made some scholars sceptical as to the prospects of the discipline:

> Thus one of the main problems with the scientific investigation of translation seems to lie in the fact that not only linguistic factors, but many other factors need to be taken into account. Since these factors belong to a variety of different areas of life, there is a question whether a comprehensive account of translation in the form of a coherent and homogeneous theory can ever be achieved. (Gutt 1991: 5)

Gutt does not seem to realize that similar doubts could be voiced by representatives of all contemporary humanities. Science in general is becoming more and more interdisciplinary.

The fact that translation studies makes use of the findings of psychology, sociology, linguistics or computer science does not mean that it is (or should be) a subdiscipline of any of them. It is definitely not a subdiscipline of applied linguistics, because language is only one of the aspects of translation.

On the other hand, different inspirations result in different schools of translation studies. It is possible to distinguish at least five "trends" in the discipline: linguistic, psychological, sociological, literary and philosophical. The best known representative of the linguistic trend is J.C. Catford (1965). More recently the school is represented by A. Bogusławski (2013), who even uses the label "translatory linguistics", and T. Krzeszowski (2016), whose approach is based on cognitive linguistics. The psychological school is represented by, inter alia, E.-A. Gutt (1991), making use of the relevance theory, and E. Tabakowska (1993). The best example of the sociological approach is the so-called descriptive translation studies with its most eminent representatives G. Toury (1995) and Th. Hermans (2007). Among numerous literary-oriented translation scholars I would like to mention E. Balcerzan (2010), A. Berman (1984), J. Brzozowski (2011). Philosophically-minded scholars deal with the so-called hermeneutics of translation, their best-known representatives being probably G. Steiner (1975) and R. Stolze (1994). All those schools compete and collaborate, all those trends overlap to some extent, since their object of study is the same – translation.

Because of the ever-growing significance and number of translations, because of the complicated and interesting nature of the mental processes involved in translating, translation studies is likely to attract more and more students and scholars and gain in importance.

References

Balcerzan, Edward. 2010. *Tłumaczenie jako "wojna światów". W kręgu translatologii i komparatystyki*. Poznań: Wydawnictwo Naukowe UAM.

Berman, Andre. 1984. *L'épreuve de l'étranger. Culture et traduction dans l'Allemagne romantique*. Paris: Gallimard.

Bogusławski, Andrzej. 2013. *Podstawy konfrontatywnej lingwistyki przekładowej*. Łask: Oficyna Wydawnicza Leksem.

Brzozowski, Jerzy. 2011. *Stanąć po stronie tłumacza. Zarys poetyki opisowej przekładu*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Catford, John C. 1965. *A Linguistic Theory of Translation*. London: Oxford University Press.

Grabe, William. 2002. "Applied Linguistics: An Emerging Discipline for the Twenty-first Century." In: Robert B. Kaplan (ed.), *The Oxford Handbook of Applied Linguistics*. Oxford: Oxford University Press, 3-12.

Gutt, Ernst-August. 1991. *Translation and Relevance. Cognition and Context*. Oxford: Blackwell.

Hermans, Theo. 2007. *The Conference of the Tongues*. Manchester: St. Jerome Publishing.

Holmes, James. 2000. "The Name and Nature of Translation Studies." In: Lawrence Venuti (ed.), *The Translation Studies Reader*. London-New York: Routledge, 172-185.

Kaplan, Robert B. (ed.) 2002. *The Oxford Handbook of Applied Linguistics*. Oxford: Oxford University Press.

Krzeszowski, Tomasz P. 2016. *The Translation Equivalence Delusion. Meaning and Translation*. Frankfurt: Peter Lang.

Richards, Ivor A. 1953. "Toward a Theory of Translating." In: A. F. Wright (ed.), *Studies in Chinese Thought*. Chicago: University of Chicago Press.

Richards, Jack C. Platt, John. Platt, Heidi. 1992. *Longman Dictionary of Language Teaching and Applied Linguistics*. Burnt Mill, Harlow: Longman.

Steiner, George. 1975. *After Babel. Aspects of Language and Translation*. Oxford: Oxford University Press.

Stolze, Radegundis. 1994. Übersetzungstheorien. Eine Einführung. Tűbingen: Narr.

Szulc, Aleksander. 1984. *Podręczny słownik językoznawstwa stosowanego. Dydaktyka języków obcych*. Warszawa: Państwowe Wydawnictwo Naukowe.

Tabakowska, Elżbieta. 1993. *Cognitive Linguistics and Poetics of Translation*. Tűbingen: Narr.

Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: Benjamins.

The New Oxford Dictionary of English. 1998. Edited by Judy Pearsall. Oxford: Clarendon Press.

LINGUISTICS

M.G. LALITH ANANDA
University of Sri Jayewardenepura,
Colombo, Sri Lanka

DOI: 10.15290/cr.2018.21.2.02

Information Structure Projects in Syntax: Evidence from Focus and Modality in Sinhala

Abstract. The major claim of this paper is that information structure related particles of Sinhala are distinct functional heads and they project in syntax. This is in line with the cartographic approach to syntax which claims that discourse related features are visible for computation (Rizzi 1997, 2004), a claim also supported by Miyagawa, (2010), and Aboh (2010), among others. The present paper seeks to validate the above claim with evidence from Sinhala, motivating the argument that discourse related features lexicalized in Sinhala drive the derivation, and these features are comparable to formal features in establishing an Agree relation.

Keywords: information structure, functional heads, Sinhala, syntax.

1. Introduction

Information structure is a term first introduced by Halliday (1967) to account for the distinctions of focus, presupposition, and propositional attitude toward entities in the discourse conveyed by phrasal intonation. It is mainly concerned with context-based information such as topic/old information versus focus/new information. According to Zimmermann and Fery (2010: 01), information structure is that cognitive domain that mediates between the modules of linguistic

competence in the narrow sense, such as syntax, phonology, and morphology, and other cognitive faculties which serve the central purpose of the fixation of belief by way of information update, pragmatic reasoning, and general inference processes.

Lambrecht (1995) defines information structure of a sentence as the formal expression of the pragmatic structuring of a proposition in a discourse. A proposition which has undergone pragmatic structuring is called a pragmatically structured proposition. According to him, the most important categories of information structure are (1) Presupposition and Assertion, which have to do with the structuring of propositions into portions which a speaker assumes an addressee already knows or does not yet know (2) Identifiability and Activation, which have to do with a speaker's assumptions about the statuses of the mental representations of discourse referents in the addressee's mind at the time of an utterance, and (3) Topic and Focus, which have to do with a speaker's assessment of the relative predictability vs. unpredictability of the relations between propositions and their elements in given discourse situations (Lambrecht 1995: 5-6). In his view, information structure is the result of the interaction between all meaning bearing levels of the grammatical system as manifested in prosody, semantics, and morpho-syntax which interact in various language specific ways. Accordingly, the sentences subjected to his analysis are of three types. (1) Constructions that express differences in the respective scope of presupposition and the assertion, differences in topic-focus structure, or differences in the cognitive status of the referents of argument expressions. All these constitute the constructions that express differences in Information Structure. As Lambrecht himself says, his study is located between both formal and functional approaches to syntax.

Irrespective of the standpoint or labelling, what is evident here is the assumption that certain formal properties of sentences cannot be fully understood extraneous to the linguistic and extralinguistic contexts that apply to those sentences. Sinhala[1] offers an interesting case in this regard with its morphological realization of information structure in the form of particles/affixes. The aim of this paper is to explore this interaction with empirical justification from Sinhala to substantiate the argument that information-structure is predetermined as such information structure related particles get computed in Syntax, rather than being added in some kind of post-syntactic component. The theoretical standpoint adopted in this paper is the cartographic approach as expounded by Rizzi (1997, 1999) and Cinque (1999). Rizzi's seminal paper on the fine structure of the left periphery (1997) expounds a proposal for decomposition of the Complementizer layer of the clause into a series of functional projections in analogy to Pollock's decomposition of the sentence eight years earlier. Motivating this decomposition by the peculiarities of complementizers of Italian and other Romance languages, Rizzi argues that interrogative and relative pronouns, topics, and foci project their own X-bar projections, and that this articulated array of projections constitutes the complementizer system (C-system). The C-system is interpreted as an interface

1 Indo-Aryan, SOV, pro-drop, spoken by the majority Sinhalese in Sri Lanka.

between two layers of an information system, one interfacing with the domain of discourse – typing the clause as interrogative, relative, adverbial, etc. – and the other interfacing with the domain of the sentence – expressing the content within IP, and determining its finiteness properties. Accordingly, the information contained in the higher structure is called the specification of Force (or Force) and the lower, more inward-looking structure headed by IP, as Finiteness.

Based on a wealth of crosslinguistic evidence, Cinque (1999) builds up the argument, that natural language clause is a construct of Moods, Modals, Tenses, and Aspects. He argues that these major clause-building categories are rigidly hierarchically ordered with respect to each other. Based on the distribution of Adverbial Phrases (AdvPs) in Italian, French, and other Romance languages, Cinque observes the presence of a head position of a functional projection to the immediate right and left of each such AdvP. Then the two independently established hierarchies, the AdvPs and the functional heads are matched systematically from left to right. The transparent semantic relation that exists between each adverb class and the contiguous head morpheme provide evidence that each AdvP is the specifier of the phrase projected by the corresponding functional head morpheme. The functional projection is considered to be structurally present in every language irrespective of the AdvP's lack of overt morphological realization corresponding to the particular functional head in the case of certain languages.

Clause structure and information structure and their interaction in both root and embedded peripheries have been central in generative grammar and have been extensively dealt with over the years with a view to understanding the properties of both and finally of UG (Kidwai 1999; Rizzi 1997, 1998; Cinque 1999; Zagona 2007; Aboh 2010; Miyagawa 2010; Ananda 2011, 2012). The extension of the X-bar schema to the functional heads –CP and TP—and the explosion of the functional domain further highlights the significance of clausal architecture and information structure in syntactic theory. Though often conceived as distinct domains, both share certain common properties that may be differently represented in root and embedded peripheries. For example, the C-domain is both an information structure domain, housing topic and focus projections as well as heads that actualize key projections of the clause – finiteness and tense. Chomsky (2005) notes that "basic tense and also tense like properties (e.g., irrealis) are determined by C (in which they are inherent: "John left" is past tense whether or not it is embedded) or by the selecting V (also inherent) or perhaps even broader context. In the lexicon, T lacks these features. T manifests the basic tense features if and only if it is selected by C (default agreement aside); if not, it is a raising (or ECM infinitival), lacking Φ features and basic tense. So it makes sense to assume that Agree—and Tense—features are inherited from C, the phase head" (Chomsky 2005: 10).

Sinhala offers fertile ground for enquiry, given its liberal use of a number of particles and lexical words to encode topic, focus, mood and modality, the consequence of which being that the morphology makes transparent the relations between information structure and clausal architecture. In addition, the role of verbal morphology in determining the particular modal, topic, focus or Wh interpretation and scope relations highlights the overt interaction of morphology and syntax in clause structure and information structure. The data for the present study consist of the

grammatical judgments of native speakers of Sinhala. Although the researcher himself is a native speaker of Sinhala, grammatical judgments of at least 10 native speakers were sought. The data presented in the following sections were first subjected to the grammatical judgments of the native speakers of Sinhala for both accuracy and verification. The analysis attempted here aligns with both the existing theoretical claims and empirical arguments.

The paper is structured in the following manner. Section 2 presents the Sinhala facts. Section 3 provides data from the embedded contexts. Section 4 presents previous research on the same. Section 5 attempts an analysis of the Sinhala facts along the particular theoretical standpoint adopted. Section concludes the paper.

2. The Sinhala facts

The canonical word order in Sinhala is SOV, though other word order variations are also possible and freely used by the speakers. Thus, SVO, OVS, VSO, VOS, OSV are other possibilities. Sinhala is a thorough-going left branching language with all type of heads including the complementizer (kiyala) occurring as the rightmost element in phrasal and clausal architecture. Sinhala is pro-drop allowing the possibility of dropping elements in both subject and object argument positions. There is no subject-verb agreement in Sinhala except for some Focus/Modal agreement where a focused element or an element under the scope of an epistemic modal marker requires the –e form of the verb (as illustrated in later sections). Nominative subjects are unmarked. Sinhala also uses dative, instrumental, and accusative subjects which are overtly marked.

Sinhala has a number of particles/suffixes to convey mood, modality, interrogative, and the information focus. As illustrated in the following sections, they can attach to any lexical category in an agglutinative fashion and take scope over the domain to the left. They can also attach at the clausal level, thereby scoping the whole clause. They are present in both root and embedded peripheries; although their distribution is not uniform in this respect, interacting as they do, with the morpho-syntax of Sinhala at different levels. For example, the presence of such a mood/modal/interrogative or information focus particle in the clause is shown in the verbal morphology in the form of an –e suffix, in the present and past tenses.

Examples (1-4) illustrate this phenomenon with respect to focus, evidential modality, and epistemic modality, although the same phenomenon occurs with respect to other modalities, interrogative, and negation, as well.

Example (1) is a neutral sentence. In (2-4), the subject, *Nimal,* is exclusively in the (narrow) scope of the focus, evidential, and the evaluative modal particles respectively.

1. *Nimal* *kaareka* *seeduw**a*** ***(Neutral statement)***
Nimal (Nom) car-def wash (Pst)
'Nimal washed the car'

2. *Nimal* *tamai* *kaareka* *seeduw**e*** ***(Focus)***

Nimal (Nom) Foc car-def wash-E (Pst)
'It was Nimal who washed the car'

3. *Nimal lu kaareka seeduw**e (Evidential modality)***
Nimal (Nom) Evid car-def wash-E (Pst)
'It is said that Nimal washed the car'

4. *Nimal ne kaareka seeduw**e (Evaluative modality)***
Nimal (Nom) Epis car-def wash-E (Pst)
'Nimal washed the car' (Speaker lacking confidence in the truth of the statement)

The same particle can attach at the clausal level, and then the whole clause comes under the (wide) scope of that particle. Example (5) illustrates this with the Evid particle.

1. *Nimal kaareka seeduw**a lu***
Nimal (Nom) car-def wash (Pst) ***Evid***
'It is said that Nimal washed the car'

Note that a crucial difference between 2-4 and 5 is that the –e suffix of the verb is absent in the latter. This differential behavior of the –e suffix highlights two things: (i) it is not simply the focus/modal particle that determines the contrastive focus/modal interpretation, but the verbal inflection also takes part in this process. (ii) It shows the scope marking potential of the focus/modal particle and the corresponding verbal morphology. That is, when the focus/modal particle attaches to any phrase level constituent, the verb inflects for –e. This creates a set of alternatives out of which one individual/entity is given saliency. But, when the same particle attaches to the whole clause, it does not inflect for the –e suffix (5). Therefore, (5) indicates that the alternative set is not available in this instance.

In both cases, however, there must be adjacency between the relevant particle and the scope marked constituent/clause. No other category (adverb etc.) other than another modal particle or focus particle can intervene between the two.

Table 1 illustrates the information structure related particles of Sinhala with examples for each.

Table 1. Information Structure related particles of Sinhala

Category	**Particle**	**Example**
MOOD EVIDENTIAL	-lu	*Nimal lu gaha kaepuw-e* *Nimal(Nom) Evid tree cut(Pst)-e* *'It is said that it was Nimal who cut the tree'*

MOOD EVALUATIVE	-ne	*Nimal ne gaha kaepuw-e* *Nimal(Nom) Eval tree cut(Pst)-e* *'Nimal cut the tree'*
MOOD EPISTEMIC	-yae	*Nimal yae gaha kaepuw-e* *Nimal(Nom) Epis tree cut(Pst)-e* *'It is doubtful that it was Nimal who cut the tree'*
MOOD EPISTEMIC: Probability/ possibility	puluwan	*Nimal gaha kapann-a puluwan* *Nimal(Nom) tree cut-Inf might* *'Nimal might cut the tree' (Epistemic possibility)*
MOOD EPISTEMIC: Probability/ possibility	vage	*Nimal gaha kapa-la vage* *Nimal(Nom) tree cut-PPt seem* *'It seems Nimal has cut the tree'*
MOOD INT(ERROGATIVE) (Q)	-da	*Nimal gaha kaepuwa-da* *Nimal(Nom) tree cut(Pst)-Int/Q* *Did Nimal cut the tree?*
MOOD CONDITIONAL	-nang	*Nimal gaha kaepu-a nang mama eya-ta baninava* *Nimal(Nom) tree cut(Pst)-a if I(Nom) he-Dat scold* *'If Nimal cut the tree I would scold him'*
COMP(LEMENTIZER)	-kiyala	*Nimal gaha kaepu-a kiyala amma kiuwa* *Nimal(Nom) tree cut(Pst)-a Comp mother said* *'Mother said that Nimal cut the tree'*
NEG(ATION)	naeha	*Nimal gaha kaepu-e naehae* *Nimal(Nom) tree cut(Pst)-e Neg* *'Nimal did not cut the tree'*
FOCUS	tamai	*Nimal tamai gaha kaepuw-e* *Nimal(Nom) Foc tree cut(Pst)-e* *'It was Nimal who cut the tree'*
FOCUS (Neg)	nemei	*Nimal nemei gaha kaepuw-e* *Nimal(Nom) Foc(Neg) tree cut(Pst)-e* *It wasn't Nimal who cut the tree*
TOPIC	-nang	*Nimal nang gaha kaepuw-a* *Nimal(Nom) Top tree cut(Pst)-a* *'As for Nimal, he cut the tree'*
MODAL (ROOT)	puluwan	*Nimal-ta gaha kapann-a puluwan* *Nimal(Dat) tree cut-Inf can* *'Nimal can cut the tree' (Root ability)*
MODAL (ROOT)	baehae	*Nimal-ta gaha kapann-a baehae* *Nimal(Dat) tree cut-Inf cannot* *'Nimal cannot cut the tree' (Root impossibility)*

IP		
VP		

Table 1 illustrates a number of significant properties of Sinhala discourse particles. Of the epistemic modals, evidential, evaluative, epistemic (except epistemic possibility), and interrogative attach to the fully inflected verb, i.e. they attach to the present, past, future, and past participle verbal forms which may be inflected for indicative/imperative/hortative /volitive/and future/irrealis moods of the verb. With respect to focus and topic, they too show a similar distribution. But in root/event modalities, the modalities of ability and permission, only the infinitive/imperative verb forms are allowed. Narrow scope marking by the modal is not possible here.

3. Embedded clauses

Topic and Focus can occur in the embedded clause. However, there are distributional restrictions. Only one topic or focus particle can occur in the embedded clause. The occurrence of one in the matrix clause and one in the embedded clause at the same time is disallowed. In the same way, multiple occurrence in the same clause is not allowed. Both Topic and Focus are incompatible with a Wh-phrase. In embedded clauses, the evidential/evaluative cannot have narrow or wide scope, thus indicating that evidentiality/evaluative modality in Sinhala is a root phenomenon. This is further supported by empirical facts as two evidential/evaluative particles (lu/ne) cannot occur in the clause simultaneously, one in the matrix and another in the embedded.

6. **Nimal [Ajith **lu/ne** horakam-karapu badu-wagayak] soyanne*
Nimal [Ajith ***EVID****/EVAL* *STOLEN-DID* *GOODS-CERTAIN] LOOK FOR-E*
It is said that Nimal is looking for certain goods stolen by Ajith

The distribution of these information structure related particles in the embedded contexts is illustrated below.

Table 2. Syntactic Properties of Focus, Topic and the Modals

Property	**Focus**	**Topic**	**Epis Modals**	**Root Modals**
Contrastive narrow scope possible				x
-e suffix on the verb in narrow scope		x		x

Clausal level scope possible		x		
Occur in root clause				
Occur in embedded clause			x	
Multiple occurrence in the same clause	x	x	x	x
One in matrix and one in embedded clause simultaneously	x	x	x	x
Compatible with a Wh in the same clause	x	x	x	x

Some insights that can be gained from the above table are that, in Sinhala, Topic and Focus behave rather differently with respect to e-marking and in narrow scope marking. Topic marking at clause level is also not possible. All information structure related particles occur in both root and embedded peripheries except the epistemic modals which do not occur in the embedded clause. All of them show similar distributional restrictions with respect to multiple occurrence and compatibility with a Wh.

4. Previous literature

Hagstrom (1998) discusses the WH question formation with respect to syntax, morphology, and semantics of Sinhala questions. He explores the interrelation between the Wh-construction and the Focus construction based on the identical distribution and scope marking properties of the Question particle "da" in Sinhala. Hagstrom maintains that the role of e-suffix is central to the understanding of the movement relation and establishing the identity of the moving particle/constituent. He proposes that e– Suffix serves a scope marking function that depends on the distribution of the Q particle. Where Q (da) is clause internal, the embedded verb is marked with –e, but a clause peripheral Q (da) does not trigger –e on the verb. He identifies a strong syntactic parallel between Wh and Focus on the basis of the above distributional evidence. He concludes that the e– morpheme is a morphological reflection of an unchecked feature and suffixation of the Q-head "da" or the focus head "tamai" can check this feature via movement. He identifies Focus as an independent head.

Heenadeerage (2002) examines the role of the e– suffix in the context of the Sinhala focus construction. He identifies three distinct types of focus in Sinhala as Constituent Focus, Predicate Focus, and Clause-Final Focus. Constituent focus corresponds to morphological marking of focus with a focus particle, where a pre-verbal constituent followed by the focus marker receives focus in the discourse. In this case the verb is e-marked. Predicate focus refers to the propositional focus where a focus particle occurs in the clause final position so that the whole proposition is focused. This does not trigger e– on the verb. The post verbal position (with the verb e-marked) where a constituent receives focus is identified as Clause Final focus. This is also identified as syntactic focus in the literature. He lists the modal particles as focus markers so that they share the same

structural position and distribution. However, Heenadeerage does not attempt a cartographic analysis as his approach is Lexical Functional Grammar.

Kariyakarawana (1998) investigates the focus phenomena of Sinhala in the theoretical framework of Government and Binding (Chomsky 1981, 1986a, 1986b) and attempts at a comprehensive analysis of the focus construction. His critical examination of focus includes the cleft construction, Wh movement, focus particles, focus and pre-supposition, and the verb marking. He lists the particles *lu* (reportative), *da* (interrogative), *ne* (tag), *tamai* (Foc) as focus markers that make any constituent immediately preceding one of them morphologically focused and observes that they attribute a contrastive meaning to the whole proposition, or a constituent that comes under the scope of such a particle thereby contributing to the propositional focus and constituent focus dichotomy. He generalizes that the different particles that encode some degree of focus and have a similar distribution are focus particles. A critical investigation of the focus/modal particles and their syntactic representation has not been attempted.

Aboh (2010) argues that information structure begins with the Numeration itself and hence is part of narrow syntax. He shows that the functional head C which carries the Wh-feature, forcing I to C movement for clause typing is in the lexicon and therefore, is part of the Numeration in a Wh-question. This shows that the lexicon contains discourse-related functional items that project in syntax. He further shows with respect to question-answer pairs, that even though speakers have the freedom of choice with regard to which linguistic expression to use in a particular context or discourse, the form of this expression is a product of syntax that directly relates to the Numeration. For example, "that the matching answer to a Wh-question contains a focused expression seems to be a requirement of the question operator in the question. Accordingly, focus assignment (which in some languages, e.g., Gunbe, require constituent displacement) satisfies a syntactic requirement generated in the question through clause-typing and Wh-licensing" (Aboh 2010: 18).

Aboh provides evidence from Maale, an SOV North Omotic language spoken in Southern Ethiopia, Lele, a Chadic language, and Gunbe to support the claim that information-structure related information is encoded in the lexicon. In Maale, all sentence types must be morphologically marked on the verb. These discourse related inflectional suffixes mark discourse modality and therefore, he argues, should be part of core syntax. In Lele, both yes-no questions and wh-questions require the presence of a sentence-final question particle. Lele displays both in-situ and ex-situ wh-questions, and the ex-situ construction occurs in the context of the focus marker *–ba*. The wh-phrase moves to the left periphery because it is focused, suggesting that Focus and Wh-movement should be separately treated as involving two probes. The important point he raises is that this particular information sensitive information should be part of the lexicon, which the speakers acquire. Gunbe offers further evidence for the same claim. In Gunbe, both topic and focus are morphologically marked (wЄ), (ya) and both occur to the right of the complementizer ɖa. Aboh suggests that these topic and focus markers are heads that project their own X-bar schema and attract the relevant constituents to Spec-Top and Spec-Foc: An argument which strongly supports Rizzi's (1997, 1998) split-C hypothesis.

Miyagawa (2010) argues with respect to Japanese that the topic/focus feature is a grammatical feature in discourse configurational languages which is equivalent to the Phi-agreement feature in agreement languages. He proposes a separate projection – aP, above TP and below CP. The *a* head may also host a grammatical feature, and when there are two grammatical features – topic and topic/focus, for example – one occurs on *a* and the other on T, and both involving A-movements. This too supports the cartographic approach to syntax, which claims that discourse related features are visible for computation.

5. Analysis

Information structure encoding in Sinhala presents a challenge to both the minimalist assumptions (Chomsky 1995) and Discourse Pragmatics/Information Structure as expounded by Lambrecht (1995). According to the minimalist standpoint, the topic/focus related information is considered a pragmatic property and hence is not well motivated in the narrow syntax. Lambrecht (1995) adopts a similar view with his pragmatic-grammatical orientation to the phenomena under study. For example, in the framework he has adopted, the definition of topic is related to the notion of subject in traditional grammar in line with the argument that the topic of a sentence is the thing which the proposition expressed by the sentence is ABOUT. Accordingly, a "topicalized" phrase may stand either in a topic relation or in a focus relation to the proposition expressed by the sentence because the first has a "predicate focus" structure and the second "argument-focus" structure. And this clear difference in pragmatic function correlates with an equally clear prosodic difference. At the level of syntax, however, the difference is not marked (Lambrecht 1995: 123).

As shown above, in Sinhala, the picture is different as focus/modality encoding takes place morphologically through particles. Essentially then these lexical items/particles should be in the lexicon before they become a Numeration, must have semantic features, and get computed in syntax. Hence, in a way, information structure of the clause is pre-determined. This indicates that what drives the derivation cannot be the formal features alone, but the feature composition of the discourse particles too. Therefore, the morphological encoding of focus/modals in Sinhala offers further empirical justification for a cartographic approach. Thus, in line with the cartographic approach adopted by Rizzi (1997) and Cinque (1999), I propose that information structure related particles in Sinhala are distinct functional heads. Their head order is determined by their order of occurrence in the clause. They are part of the Numeration and that they project in syntax. Also, as it will be shown below, a Wh-operator cannot be focused (07), and Focus and a Wh do not co-occur in Sinhala (08), prompting the argument that there is only one head position for both which attracts the relevant constituent to its Specifier.

7) **Kauda tamai gaha kaepuw-**e**?*
Who FOC TREE CUT-E
'It is who cut the tree?'

Also, a Wh and a focused constituent are not compatible in the same clause showing that there is only one landing site for both in the root clause, as in (8).

8) **Kauda gaha tamai kaepuwe?*
Who tree FOC CUT-E
'Who cut the tree' (tree focused)

This holds in the embedded contexts too, as a Wh and a focused constituent are not compatible simultaneously in the embedded clause either, as in (9).

9) ****Sunil*** *[Kauda gaha tamai kaepuwe kiyala] aehuwa?*
Sunil [Who tree FOC CUT-E COMP] ASKED?
'Sunil asked who cut the tree' (tree focused)

Now, we must find out the structural position of the epistemic modals with respect to other information structure related particles of Sinhala. Taking examples from a wealth of languages, Cinque (1999) proposes that epistemic modals are located higher than root modals (higher than Tense as well) so that the former has scope over the latter. His hierarchy of functional heads shows that epistemic modals are outside the scope of Tense but within the scope of evaluation time specified in CP (ForceP). This line of argument is also in line with Stowell (2004). Stowell shows that epistemic modals are construed in relation to the evaluation time of their clause. Stowell concludes that epistemics can have both past and present forms but are associated with the evaluation time of the clause. In line with Cinque (1999), and Stowell (2004) I propose that the epistemic Modals in Sinhala are located in the C- domain, below Force. The evidence for the above claim can be presented as follows.

10) *Nimal thamai lu/ne gaha kaepuwe
Nimal Foc Evid/Eval tree cut(past)
'It is Nimal /as people say it is Nimal who cut the tree'

11) *Nimal gaha kaepuwa lu/ne thamai
Nimal tree cut(past) Evid/Eval Foc
'Nimal cut the tree as people say / indeed'

The examples show that both Focus and Epistemic modal particles cannot co-occur, either in narrow scope marking or in broad scope marking. This further indicates that both Focus and Epistemic Modals compete for the same Head position.

Now let us further examine in what ways focus/modality/Wh/Q in Sinhala interacts with the verbal system. We noted in the preceding sections that in narrow scope marking of the focus/epistemic modal/Wh, the verb ends in –e form as opposed to neutral/declarative –a form (12-17)

*12) Nimal lu gaha kaepuw-e (*kaepuw-a)*
*Nimal Evid tree cut(Pst-E) (*Past)*
'It is Nimal, as they say, the one who cut the tree'

*13) Nimal tamai gaha kaepuw-e (*kaepuw-a)*
*Nimal Foc tree cut(Pst-E) (*Past)*
'It is Nimal who cut the tree'

*14) Kauda gaha kaepuw-e (*kaepuw-a)*
*Who tree cut(Pst-E) (*Past)*
'Who cut the tree?'

*15) Kauda gaha kaepuw-**a***
*Somebody tree cut(Pst-**a**)*
'Somebody cut the tree?'

*16) Nimal da gaha kaepuw-**e***
*Nimal (Nom) Q tree cut(Pst-**e**)*
'Was it Nimal who cut the tree?'

*17) *Nimal da gaha kaepuw-**a***
*Nimal (Nom) Q tree cut(Pst-**a**)*
'Was it Nimal who cut the tree?'

Note that in (14-15), the –e suffix is crucial for Wh force/interpretation, without which the Wh word simply becomes an existential pronoun (somebody).

One notable feature of Sinhala is its lack of Agreement. The verb inflects for Tense in Sinhala (example 18). However, the verb does not inflect for person/number/gender agreement (Phi-agreement) (example 19).

*18) Nimal kaareka soodanava/seeduw**a***
Nimal (Nom) car-def wash (Prs)/wash(Pst)
'Nimal is washing the car/Nimal washed the car'

*19) Nimal/mama/api kaareka soodanava/seeduw**a***

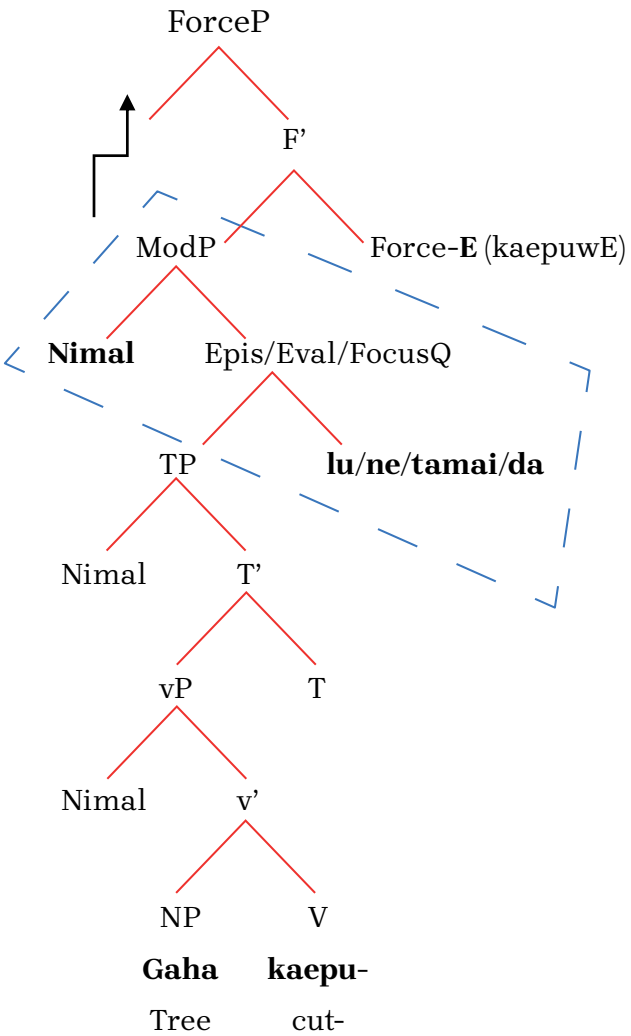
Nimal/l/ We car-def wash (Prs)/wash(Pst)

'Nimal/l/We are washing the car'

'Nimal/l/We washed the car'

However, the fact that the verb inflects for the –e form (soodannE/seeduwE) when there is a modal/focus/Q/Wh particle in the clause having narrow scope indicates some form of agreement. I propose that this constitutes modal/focus agreement in Sinhala with a ModalP which has features of Wh, Focus and Modal where all are in complementary distribution with each other. This claim is also in line with Miyagawa (2010) who motivates the argument that topic/focus features are computationally equivalent to Phi-features and trigger agree relations. I propose that a DP moves to the Spec of the ModP triggering Spec-Head agreement. And then this whole ModP moves to Spec ForceP to agree with E suffix of the verb (kaepuw-e as opposed to neutral kaepuw-a) which also marks the illocutionary Force of the utterance (20).

20)



6. Conclusion

This paper presented the argument that discourse related features lexicalized in Sinhala drive the derivation, and these features are comparable to formal features in establishing an Agree relation. This is in line with the cartographic approach to syntax which claims that discourse related features are visible for computation (Rizzi 1997, 2004), a claim also supported by Miyagawa (2010), and Aboh (2010) among others. It was shown that in Sinhala, Wh-, Focus, Evidential Modality, and Epistemic Modality are morphologically realized in the form of particles suffixed to a constituent. In such cases, a verbal argument or adjunct can come under the scope of the Focus or Modal particle. When such a particle marks narrow scope, the verb should take the –e ending, as opposed to the neutral –a ending. I proposed that these information structure related particles are Functional Heads carrying the relevant feature, they are in the lexicon, and that they become part of the Numeration and they project in syntax. Since Focus, Modal and Wh- do not co-occur in Sinhala, there is only one projection (ModP) for all which attracts the relevant constituent to its Spec. I considered the –e suffix as a reflex of a discourse Agree relation though in Sinhala there is no Phi-agreement.

References

Aboh, Enoch O., Hartmann, K., Zimmermann, M. (eds.), 2007. *Focus Strategies in African Languages*. Berlin. New York: Mouton de Gruyter.

Aboh, Enoch, O. 2010. Information Structuring Begins with the Numeration. *Iberia* vol 2.1, 12-42.

Ananda, L. 2011. *The Focus Construction in Sinhala*. Germany: Lap Lambert Publishing.

Ananda, L. 2012. *Clausal Complementation in Sinhala*. Unpublished PhD thesis. Jawaharlal Nehru University.

Chafe, W.L. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View". In: CN. Li (ed.), *Subject and Topic*, 25-55. Associated Press, New York.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. 1981. *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. 1986. *Barriers*. Cambridge, Mass: MIT Press.

Cinque, G. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective.* Oxford: Oxford University Press.

Cinque, G. & Rizzi, L. 2008. *The cartography of syntactic structures*. CISCL Working Papers.

Hagstrom, P. A. 1993. *Decomposing questions.* Doctoral dissertation. MIT.

Henadeerage, D.K. 2002. *Topics in Sinhala syntax.* Doctoral dissertation. Australian National University.

Halliday, M. A. K. 1967. Notes on Transitivity and Theme in English. *Journal of Linguistics* 3, 199-244.

Kariyakarawana, S. M. 1998. *The Syntax of Focus and WH Questions in Sinhala*. Colombo: Karunaratne and Sons LTD.

Kidwai, A. 1999. "Word order and focus positions in Universal Grammar". In: G. Rebuschi & L. Tuller (eds.), *The grammar of focus*. 213-244. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lambrecht, K. 1995. *Information structure and sentence form*. Cambridge: Cambridge University Press.

Miyagawa, S. 2010. *Why Agree? Why Move*? Cambridge: The MIT Press.

Rizzi, L. 1997. The fine structure of the left periphery. In: L. Haegeman (ed.), *Elements of Grammar.* 281-338. Dordrecht: Kluwer Academic Publishers.

Rizzi, L. 1998. *On the Position of Int(errogative) in the Left Periphery of the Clause*. Manuscript. University of Siena.

Vallduví, E. 1992. *The Informational Component*. Garland: New York.

Zagona, K. 2007. "On the syntactic features of epistemic and root modals". In: Luis Eguren & Olga Fernandez (eds.), *Co-reference, Modality and Focus.* 221-236. Amsterdam/Philadelphia: John Benjamins.

Zimmermann, M., Fery, C. 2010. *Information Structure*. Oxford: Oxford University Press.

PAWEŁ DZIEDZIUL
University of Białystok

DOI: 10.15290/cr.2018.21.2.03

Contronymy and Semantic Primes

Abstract: Contronymy, that is sense opposition invoked by one word, can pose a serious conundrum from a theoretical standpoint. Nonetheless, the prime concern of this paper is to introduce the phenomenon into a broader discussion within theoretical linguistics. To be more specific, the question at hand is: what kind of comprehensive and coherent theoretical construct can be adequate for semantic representation of contronymy? It will be demonstrated that the particular sense opposition can be classified as being linked with direct negation. A theoretical vantage point will be presented that addresses the cause of opposition via the means of the natural semantic metalanguage theory. This approach may shed some light on how to deal with the problem from a cognitive perspective. The underlying methodological assumptions of the presented framework, based on the idea of semantic primes, prove to be a coherent tool for encapsulating radical sense opposition manifested by contronyms. As an addendum to this prolegomena there will also be presented a brief discussion of some of the implications of contronymy for fields such as the theory of the human mind, natural language processing, artificial intelligence, machine translations and big data structures.

Keywords: contronymy, sense opposition, primitive concepts, semantic primes, negation, natural semantic metalanguage theory.

What is frequently assumed about natural language is that it is to a fair degree different from formal languages and formal systems in general. On an interesting opening note, as Wolniewicz (1980: 7) pointed out, it can be stated that the whole opposition of natural language vs. artificial/constructed/formal language is somewhat skewed, as the modifier 'natural' in the latter may imply that symbolic logic and mathematics are somehow less natural, or at least man-made. This, however, is not the case, as logical laws and their imperatives seem no less natural, in the sense that they are discovered rather than designed. This can be considered as an aspect that is shared by both natural and formal languages alike. Another example of the misunderstanding is the assumption that formal languages are fully explicit whereas natural languages simply are not. Yet, for example in the "language" of mathematics, polysemy is not uncommon. The symbol of minus may mean the operation of subtraction or indicate a negative number. Although the two ideas are linked at least intuitively, and the syntax of a mathematical equation resolves the ambiguity, the polysemy is quite obvious. A similar idea is present in programming languages under the term *operator overloading*, where programmers can reassign the semantics of an operator based on certain rules. Even so, what one would expect form a formal system is that one sign/word of that system would

not convey opposite meanings, especially within a syntactic context. This is, however, the case when it comes to natural languages. The phenomenon of *contronymy*, which is the prime concern of this paper, is based on the idea that two opposite senses may be evoked by one word. To give an example one can examine the verb 'to seed' and come to the conclusion that it can be used in two opposite meanings: 1) as the action of putting seeds into some sort of object, or 2) as an action that is based on removing seeds. Therefore, such an instance may be used to show a strong opposition between human languages and formal languages(or formal systems in general), as contronyms, as will be shown on further pages, also preserve their particular polysemy within well-formed phrases and sentences. To show what properties a mathematical equivalent would have to possess to mimic entities like contronyms, one may consider a minus sign, '-', which can be used to symbolize not only subtraction, but also an idea that would contradict subtraction, such as addition, and moreover all of that within a full blown equation. Such a situation seems to be uncommon to say the least within arithmetic or formal/programming languages, and it is hardly imaginable that such singularity would be of any use.

The existence of the phenomenon of contronymy leads to questions about how distant natural languages are from formal languages, and whether formal methods are suitable for the analysis of English, Polish, or any ethnic language. The idea of such stark sense opposition at first may seem difficult to capture from a strict theoretical perspective. Nonetheless, as will be demonstrated, contronymy may be conceptually linked to negation, and by this token can be summarized and brought together using precise terminology. The main thread depicted here will lead to conclusions on what type of theoretical approach may be adequate for the natural language, and also what kind of impact contronymy may have on disciplines other than theoretical linguistics that would demand general cohesion.

To briefly show the significance of this phenomenon in a contemporary context, suffice it to say that within big data structures it is not uncommon to have contradictory objects. Computer software that has as its very purpose processing such structures, or even inferring on the basis of such constructs (automated reasoning), has to somehow cope with the inconsistencies. Contronyms will be presented as an example of such a predicament. Any linguistic system that holds in its dictionary entries like the verb 'to seed' has to manifest serious consistency problems.

The paper will be concluded with a brief illustration of a theoretical approach to contronyms that seems viable. Natural semantic metalanguage theory (NSM), developed and promulgated by cognitive linguists Anna Wierzbicka and Cliff Goddard, considering its particular characteristics that will be depicted in the last but one section, is conceptually compelling. Semantic primes employed by the discussed theory include the logical operator of negation in the form of the concept NOT. With the use of this elemental particle of meaning it is possible to provide explications of contronyms using the natural semantic metalanguage, an approach that by definition does not use signs/words that evoke opposite senses. Additionally, the discussed framework of NSM theory enables us to look at the human mind as a coherent system, despite the existence of words that can be summed up as examples of antilogy.

Contronymy

Within the discipline of linguistics in recent years the phenomenon of two opposite senses concealed under one lexical unit has been examined to some extent, although only from the descriptive perspective (see Karaman 2008). As Karaman puts it: "That such a phenomenon exists in natural language is, according to some researchers, unthinkable since it seems to be an impossibility that a linguistic sign could signify ideas which contradict each other" (Karaman 2008: 173). In light of those words a question comes to mind: what type of theoretical construct can account for the contradictions embedded in one single word in a coherent way? Yet, before presenting a systemic approach to the discussed singularity, the general characteristics of the phenomenon have to be brought forward.

Words that possess opposite senses are labelled in the literature under terms such as: 'contronyms', 'antilogies', 'auto-antonyms', 'two-faced words', 'Janus words' and many more. All those terms are treated as synonyms, although there are substantial differences among members of this category. Karaman (2008: 175) proposes labelling the whole phenomenon under the umbrella term 'contronymy' and dividing it into subcategories that would account for the alterations, such as contronymy based on antonymy, incompatibility, reversivity, complementarity and conversivity. This is due to a broad definition that groups all members that have the discussed sense opposition which states as follows: "Contronymy is a form of polysemy that can be defined as sense opposition at the micro-level. This occurs when a minimum of two senses of a polysemous lexical item contradict each other" (Karaman 2008: 175).

For the purposes of this paper the presentation will be limited to a few subcategories. At interest here are sense oppositions, such as those based on complementarity, incompatibility, as well as reversivity, although it can be claimed that the results are equally adequate for all contronyms[1].

The first example is a contronym based on reversivity in the form of the verb 'skinned', which can be used in opposite senses:

(1) John *skinned* the banana.

(2) John *skinned* the sofa.

The polysemy based on reversivity, defined here, however, intuitively, is apparent and is based on the idea of a process that has two directions, and spans two opposite terminal states (Karaman 2008: 184, 185). In the first sentence the meaning of the phrase revolves around *removing* the skin of the banana, whereas in the second one it is *applying* the skin that is relevant. Therefore, the verb can be used in two opposite ways: (S1) to express the removal of something, and (S2) to express applying something. To add to the general descriptive analysis conducted by Karaman, I would like to signify that one meaning also indicates the negation of the other. The acceptance of the sense (S1) implies

1 For a broader presentation of typology of contronyms, also with regard to antonymy, see Karaman (2008) and Klégr (2013).

that the negation of (S2) is true, as applying something means *not* removing the very same thing, and *vice versa*. When it comes to sentence (2), however, it is possible to imagine that somebody is actually removing the skin from a leather sofa. Therefore, for example in sentence (2), it is possible that the same sentence may describe two actions simultaneously where one implies the negation of the other.

The deeper idiosyncrasy of this verb can be brought forward by analysing it in the form of a predicate-argument structure. If the verb 'skinned' is treated as a monotransitive verb, as in the above examples (1) and (2), both senses have the same (as it is called within linguistics) valency, with the positions for the subjects and direct objects filled accordingly. Increasing the number of arguments and treating it as a ditransitive verb, by for example introducing some sort of instrument in the form of a prepositional object, will not remove the discussed opposition. Therefore, there is little to say about the nature of this stark polysemy by analysing the structure of the sentence, or to be more precise, when it comes to the argument structure of the verb. This is especially evident while imagining that one sentence can be understood in two opposite ways. Hence, approaches that are based upon the predicate-argument structure will suffer in situations such as the one described above. It has to be stressed that such methods are prevalent among cognitive linguistics and are frequently employed in natural language processing as well as artificial intelligence: theories such as case theory (Fillmore 1968), frame semantics (Fillmore 1976) or construction grammar (Goldberg 1995) to name a few. It is worth noting that the combination of the two latter theories was already implemented in natural language processing and reasoning (see Minsky 1975 and Spranger, Suchan, Bhatt 2016). In this sense the idea of a *frame* can provide the necessary background of non-linguistic knowledge that is relevant for interpreting the language structure (what kind of instrument is proper for removing and what kind of instrument is proper for applying skin to an object, etc.). In this case the ambiguous structure is filled by praxeological knowledge. The same scenario and analysis applies to other verbs that exhibit opposition to remove/deprive of vs. to add/yield, just like the verb 'to seed' mentioned in the beginning of the article.

Another interesting example is the word 'screen'. It may be considered as an example of a contronym based on incompatibility. This category is defined by Karaman (2008: 178) as "terms which denote classes sharing no members. This lexical element can be used regarding things that have opposite functions". The lexeme 'screen' can therefore reference an object that allows a certain observer to see a particular phenomenon, just as in the sentence below:

(3) The phenomenon *y* can be seen because of the *screen*.

Simultaneously, its opposite meaning can be detected when an object shields the observer from a phenomenon:

(4) The phenomenon *y* cannot be seen because of the *screen*.

Therefore, 'screen' may refer to things that have the function of showing something (television screen), or of shielding from something (smoke screen). Within the English language the difference of meaning cannot be mapped to the morphological structure. The root of the word 'screen' creates inflections such as 'screened', where the discussed singularity also persists.

(5) The phenomenon *y* can be seen because it is *screened*.

(6) The phenomenon *y* cannot be seen because it is *screened*.

A slightly different situation however may be observed if the word 'screen' in sentence (5) would be supplemented by a different preposition, as in 'The phenomenon y can be seen on the screen'. Therefore, syntax can be at work when it comes to the disambiguation of this particular unit. On the other hand, if there is insufficiency of string structure, the ambiguity arises, as in:

(7) *Y* is *screened*.

In this case there is clear ambiguity in the form of two potential opposite senses, and yet again the first meaning implies the negation of the other, as showing something by the means of something means not concealing something using the same thing.

The next few examples are based on mutually exclusive binary pairs (binary opposition). Examples of such contronyms (in the sense of imposing contradiction) can be shown by employing negative affixation to certain verbs or adjectives. For instance, English safety labels often explicitly enumerate together 'flammable' and 'inflammable' warnings on combustible materials, due to the fact that the prefix 'in-' may be incorrectly interpreted as changing the meaning of 'flammable' to 'not flammable'. Such lexical entities have the meaning that can be summarised with the use of symbolic logic as x and $\neg x$.

Another such example, one that again might be significant when it comes to machine translations, is the contronym 'periodic' with its low-profile ambiguity. Consider the following sentence coming from Karaman (2008: 181):

(8) *Periodic* outbreaks of the disease are inevitable.

Depending on how accustomed the reader is with the word 'periodic' he can interpret the instance of an outbreak either as happening regularly (S1), or in an irregular manner (S2). Therefore, the relation between two senses is contradictory. Again, the discussed contronym bares the character of direct negation. Only after short pondering does it become hard to believe that outbreaks of disease happen at stable intervals. However, this cannot be treated as an inference based on linguistic knowledge.

The final example again comes from Karaman (2008: 183). At the micro-level, contronymy of complementary is the most extreme type of sense-opposition, and this time the preposition 'around' can be used in two distinct senses:

(9) S1: My beta fish is swimming *around* the bowl.

(10) S2: We sat *around* the table.

As Karaman (2008: 183, 184) elaborates: "In the above examples we can observe a clear boundary between the two senses. In S1, around can be associated with 'the satellite object is part of the pivotal object' (i.e. 'the satellite object is in the interior of the pivotal object'), whereas in S2 it can be associated with 'the satellite object is not part of the pivotal object' (i.e. 'the satellite object is on the exterior of the pivotal object')". This instance illustrates the idiosyncratic character of the natural language in a very clear way. Contronymy of complementarity is an example of stark binary sense opposition. The author also cites a more mathematical explication of the nature of this type:

> Within aspect A there are the lexical elements α1, α2, ... α1 is complementary with α2 if from the existence of α1 the non-existence of α2, from the existence of α2 the non-existence of α1, from the non-existence of α1 the existence of α2, and from the non-existence of α2 the existence of a1 can be concluded." (Lutzeier 2007, quoted after Karaman 2008: 183)

Again, the discussed phenomenon manifests itself in the form of negation, as for the word 'around' if S1 is true, it implies that ¬S2 is true, and *vice versa*.

As shown, the discussed sense opposition is closely linked to negation. This is due to the fact that one contronym possesses two senses that are related in such a way that one is, or implies, the negation of the other. It is easy to recognize that within one database entries like 'screen', 'periodic', or 'around', in the way that they function in the natural language, have to lead to conflicts. It is also apparent that the risk of the logical error of equivocation, that is the situation when one term is used in two opposite senses, is therefore substantial. It has to be stressed, however, that the fact that those words can have highly fluctuating meanings (even from one extreme to the other) does not mean that those words can mean anything, or that in linguistics "everything goes". The range of possible senses of a word still has a limited scope that can be precisely accounted for from the perspective of cognitive linguistics.

Contronyms and other sense relations and features

Contronymy is also an interesting topic from the perspective of other sense relations and their features. The phenomenon was analyzed in this manner by Klégr (2013) under the label of enantiosemy. The definition of the term used in the mentioned article is: "A case of *polysemy in which one sense is in some respect the opposite of another" (Matthews 1997: 122). Klégr gives an elaborate explanation on exactly why this sense relation should be classified within the realm of poly-

semy. First of all, there are at least two meanings, therefore monosemy is excluded. The two relevant senses are without a doubt related, therefore homonymy is also not applicable. Additionally, it cannot be said that vagueness is a feature of those meanings. The possible interpretations that fall under the scope of a contronym are clear. Hyponymy and hypernymy, although they are also polysemous categories that can have multiple word-internal sense relations (for example 'pear' understood as a tree or a fruit), also logically do not seem plausible comparans, not to mention synonymy.

Antonymy, on the other hand, seems the most intuitively related semantic concept to contronyms. Some of the types of relations are indeed shared by antonyms and contronyms (directionality, converseness, etc., see Karaman 2008; Klégr 2013). Although the phenomenon of contronymy is word-internal compared to the typical word-external opposites, some analysis can be carried out on both levels. This is to say that typical antonymy is a sense relation between two distinct and yet opposite words like hot/cold, whereas contronymy has the same opposite sense relation, but in reference to just one word. Therefore, the link between the two is there. As Klégr (2013: 19) puts it: "The fact that the same set of relations operate between and within lexical items is no doubt of cognitive significance". The question that rises is whether, and if so to what extent, contronymy can be subjected to similar analysis as antonymy. Markedness would be a feature that comes to mind of antonymous pairs that could be applied. Yet again, it is another example of the difficulties that one encounters during examination of contronymy. Unlike in antonymy over two distinct words (hot/cold), or with the use of morphological negation (happy/unhappy, honest/dishonest etc.), there is no difference in the form of the two meanings. Therefore, the apparent differences between antonymy and contronymy may have an impact on the applicability of tests for markedness. In fact it seems impossible to examine whether one sense may be dominant or broader than the other when it comes to contronyms. Lehrer (1985: 398) gave an example of a useful tool that allows the identification of a marked member of the antonymy pair. It has to be noted that here antonymy is understood in the narrow, technical sense as an example of a sense opposition that has a gradable scale between the two extremes. Let's consider the questions 'How happy are you?' and 'How unhappy are you?', which reveal that the negative form carries an additional meaning that is not present in the first. This is to say that the negative form implies that the person asked is actually unhappy, and the question is just about the level of his unhappiness. The positive form does not hold this supposition (the marked member of the relation is neutralized). Sentences like 'How *custom* was the service?' ('custom' understood as a common practice, or a special treatment) seem not viable for such analysis.

On the other hand, Lehrer (1985: 399) provides a useful way of identifying the marked member of the antonym pair that can be used for contronyms. This can be done with the use of sentences that discuss proportions and ratios, as in:

John is { twice / half } as tall as Bill.

*Sally is { twice / half } as short as Sue.

By applying the same device to the contronym 'fast' (referring to either something moving quickly or something solid and unable to move) as in the sentence: 'John is twice/half as *fast* as Bill' only one meaning is preserved, and that is the one revolving around quick movement[2]. Therefore, features of antonymy such as markedness can be, at least partially, subject to similar analysis as antonymy.

On a similar note, in some cases the meaning of a contronym can be limited to only one of the possible senses by concatenating a negative prefix. For example, words like 'to unbolt' (with the opposition of 'to bolt' as in to secure/to flee) or 'to unbuckle'('buckle' - connect, or break or collapse) have only one meaning.

One of the meanings of the pair of opposites can also bear more distinctive features over the other. An example of such a case would be the verb 'to skin', already discussed in the previous section, where one of the senses may exhibit a different valency in special circumstances, and in that sense can be grammatically marked. Klégr (2013: 14) provides additional information that supplements the analysis conducted earlier on the verb 'to skin':

> In some (…) verbs enantiosemy is associated with in/transitivity; the intransitive meaning of the *cow milks / the wound soon skinned* is connected with the absence of the object due to the incompatibility of the subject with the agentive role. The opposite meaning is associated with transitivity: this use primarily results from the meaning of the verb which determines what can function as the object.

Another instance would be the derivation of the verb 'to screen', also mentioned previously. In the Polish language this derivation possesses only one sense (the verb can only be used when something is used to block some sort of phenomenon). Additionally, the use of this verb is rather technical, and is therefore significantly marked by frequency.

Another possible emanation of contronymy that should be mentioned is the use of two opposite senses under one lexical unit as a rhetorical device or as a colloquial, euphemistic or ironic expression. An example of such a type would be the use of the word 'amazing' when something went wrong, or using the phrase 'you're bad' in a situation when somebody achieved success. Klégr comments on this phenomenon in the following words:

> Nevertheless, there is a fundamental difference between a rhetorical figure of speech and antonymous polysemy, a trope being a contextual use based on ad hoc interaction which expresses the intended meaning by an implicature, while enantiosemy presupposes the existence of two distinct lexical units. (Klégr 2013: 12).

In light of the above comment, the rhetoric/ironic devices were and will be excluded from further analysis in this paper. For deeper analysis see (Klégr 2013: 15, 16).

---

2 It has to be noted that this example of a contronym does not manifest a gradable scale compared to the word 'custom'. For more details on markedness and antonymy see Lehrer (1985).

Natural Semantic Metalanguage Theory (NSM)

As mentioned earlier, the topic of contronymy has not yet been addressed via a "global" theory of language. Therefore, the main aim of this paper is to employ a linguistic theory that has gained some recognition in the literature to account for the discussed extreme sense opposition, namely Natural Semantic Metalanguage theory (NSM). It can serve as an illustration of how the phenomenon can be approached from the cognitive linguistics standpoint. The semantic primes/primitive concepts that this theory employs may be used to explicate the meaning of contronyms without falling into contradictions. This is due to the fact that the concepts of the natural semantic metalanguage have, by definition, only one distinct meaning. In essence NSM theory defines the human mind as a system that is based on coherent and consistent units of meaning and their combinations. The dictionary of the natural semantic metalanguage, although it may express all the meanings of any ethnic language, in essence consists of fewer words than any ethnic language. The multiplicity of meanings, including those that are contradictory, can be formed by combining the meanings of the semantic primes that are in core non-contradictory. The language of thought that is composed of the elementary particles of meaning, or *lingua mentalis* as Wierzbicka would put it (see Wierzbicka 1980), in the light of NMS theory is a consistent and explicit structure in that particular sense. As the proponents of this approach claim, this is a comprehensive approach to the semantics of the natural language (see Goddard 2003), and therefore the use of this theory may be considered as an expansion on what already has been done within the analysis of contronymy.

Just to give a brief outline of this framework, suffice it to say that it is a principle of *tertium comparationis* for the juxtaposition of different ethnical languages. Because of the fact that this theory uses primitive concepts that are supposed to be innate, it is a useful tool to circumvent the problems of ethnocentricity, and also a few other difficulties in constructing lexicons such: *circulus in definiendo,* and *regressus ad infinitum* (see Wierzbicka 1996: 11; Goddard and Wierzbicka 2001: 183). The particular characteristics of the framework discussed here can also help to solve other issues such as those raised by the existence of contronyms.

Anna Wierzbicka associates the main foundations of this theory with Leibniz's concept of the "alphabet of human thought". This is to say that all languages contain a palette of elementary units of meaning that are so clear and intuitive that they do not require any definition (Wierzbicka 1996: 11). For instance, the particle SEE, just as the particle KNOW (which was apparent to Descartes) are so elementary for human cognition that to give a satisfactory definition of this unit is unthinkable (Wierzbicka 1996: 48). Therefore, NSM theory postulates the existence of a finite set of universal concepts that is an intersection over the sets of all concepts of all ethnical languages. Moreover, this pallet of semantic primes is stable, native among human beings, and its members may be combined to form *more geometrico* complex meanings.

One of the more recent accounts on the set of natural semantic metalanguage units includes particles such as (Goddard 2012: 712-715):

Table 1. The set of NSM units

I, YOU, SOMEONE, SOMETHING~THING, PEOPLE, BODY	substantives
KIND, PART	relational substantives
THIS, THE SAME, OTHER~ELSE	determiners
ONE, TWO, MUCH~MANY, LITTLE~FEW, SOME, ALL	quantifiers
GOOD, BAD	evaluators
BIG, SMALL	descriptors
KNOW, THINK, WANT, FEEL, SEE, HEAR	mental predicates
SAY, WORDS, TRUE	speech
DO, HAPPEN, MOVE, TOUCH	actions, events, movement, contact
BE (SOMEWHERE), THERE IS, HAVE (SOMETHING), BE (SOMEONE/SOMETHING)	location, existence, possession, specification
LIVE, DIE	life and death
WHEN~TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORTTIME, FOR SOME TIME, MOMENT	time
WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE,INSIDE	space
NOT, MAYBE, CAN, BECAUSE, IF	logical concepts
VERY, MORE	intensifier, augmentor
LIKE~AS~WAY	similarity

The sets of elementary units of meaning undergo evolution. However, according to leading researchers within the field, the power of this set should not exceed 100 elements (Goddard and Wierzbicka 2001: 185).

For reasons of limited space not all units can be discussed here, and the remarks will be limited to the most important unit of NSM from the perspective of this article – namely the particle NOT. The concept of negation was not included in the early literature concerning natural semantic metalanguage theory (cf. Wierzbicka 1980). Initially, Wierzbicka thought that although all the evidence pointed to the fact that the concept of negation is universal, meanings like 'don't want', or the idea of 'diswant' should not be considered cognitively primal and primitive. Wierzbicka initially sought to tie the meaning of the unit with the unit WANT, as that would solve many problems regarding the disproportions between the meaning of 'want' and 'don't want' and 'know' and 'don't know' ('don't want' may mean that somebody does not want something that much, and does not have to necessarily mean that he/she does not want that something at all). However, this approach proved to be unsatisfactory in the long run. Finally, the factor that tipped the scales was the evidence on how quickly and adequately children in the early stages of language acquisition use concepts related to direct negation (see Wierzbicka 1996: 89-91). Data pointed to the interpretation that NOT is used in a discreet and combinatorial way in the sense that it is separated from other concepts.

To recapitulate, in essence, the particles of the natural semantic metalanguage can be understood similarly to the prime numbers in mathematics. That is the numbers that are members of the natural numbers category greater than 1, and cannot be expressed by multiplying two smaller natural numbers. And conversely, a finite sequence of the products of prime numbers may be used to form any natural number that is not a prime number. Therefore, a natural number greater than 1 that is not prime is called a composite number. Moreover, every natural number that is not a prime number can be factorized as a product of primes. Consequently, prime numbers can be regarded as equivalents of the primitive particles of meaning used in NSM theory, and composite numbers as compound concepts. The aim of this section is to demonstrate that those atoms of meaning may also be used to explicate contradictory words present in particular ethic languages, and this is where the analogy between mathematics and natural language breaks down.

Within NSM theory, contronyms regain their clear and intuitive interpretation. This is due to the fact that by definition the units of NSM are the elementary particles of meaning, which means that they possess only one meaning, and in that sense the dictionary of this theory is consistent. The most important particle of the NSM set for expressing the sense opposition is the logical concept NOT. This cognitive analogue of the logical operator of negation positioned within the definition of a lexical unit can account for sense opposition as well as contradictions congenital in contronyms. For example, in regard to the word 'screen' in the perspective of NSM, it is possible to denote a shared set of universal particles as well as its feature that accounts for the difference in two discussed senses. That is the substantive SOMETHING, mental predicate SEE, non-mental predicate IS, and the imagination and possibility unit IF… WOULD. The difference between the

two senses comes from the innerplacement of the additional shared unit, the logical concept NOT that is contained in one of the possible interpretations of the word 'screen'. In this chain of reasoning it is possible to reconstruct two *compound concepts* of the lexical unit that are as follows:

D_{screen} =

a. *x* is something

b. *y* does see *z*

c. if there is no *x*, *y* does not see *z*

$D_{screen'}$ =

a. *x* is something

b. *y* does not see|hear *z*

c. if there is no *x*, *y* does see|hear *z*

Although the senses of the two compound concepts are based on an antilogy, in the metalanguage they contain not only a common set of NSM units, but also, to an extent, a common set of strings of NSM units. The difference in meaning comes from two universal particles, NOT and SEE, that form a relation that is the source of polysemy. Consequently, two of those senses, within the English language, may connect to different syntactic structures regarding the prepositions 'on' and 'because of'. Interestingly, for the definition to be complete another unit has to be added, namely the particle HEAR. However, it can only supplement the second meaning of the word 'screen' that stands for an obstruction of sound. In other words, when it comes to auditory functions of the human body the word 'screen' can represent only one meaning. Nonetheless, the particle NOT within the natural semantic metalanguage may serve to reproduce the sense oppositions of the discussed contronym.

Another interesting example is the contronym 'periodic', as its sense opposition is based on the contradiction between the concepts of 'regularity' and 'irregularity', or 'not regularity'. The definitions are as follows:

$D_{periodic}$ =

a. this happens

b. if it does not happen for some time, it will happen

c. it does not happen once or twice, but many

d. if it happens now after a long time, then after that it will happen after a long time

e. if it happens now after a short time, then after that it will happen after a short time

$D_{periodic'}$ =

a. this happens

a. if it does not happen for some time, it will happen

a. it does not happen once or twice, but many

a. if it happens now after a long time, then after that it will not happen after a long time

a. if it happens now after a short time, then after that it will not happen after a short time

As depicted by the above explications the sense opposition can again be brought together by negation. Component (a) indicates that it is an event, (b) that it is not instantaneous (duration is involved),and it happens repeatedly, (c) that it happens many times, and both (d) and (e) define whether those events happen in the same increments of time or not.

The preposition 'around' can also be analysed in this framework, although its meaning is slightly more complicated. This is due to the fact that to express the idea of roundness it is necessary to use the so called *semantic molecule*. While the investigations conducted within the framework of NSM theory expanded, it began to be apparent that plausible explications of many of the elements of the lexicon require the usage of more complex structures than just the primes and universals (Goddard 2012: 10). This is where the notion of semantic molecules comes into play. The members of this category are still divisible into combinations of primes of NSM and may be used as stepping stones when direct explication by the atoms of meaning is not possible. Nonetheless, the two opposite senses of the discussed polysemy of different types can still be reproduced with unequivocal and unambiguous primes of natural semantic metalanguage, and the differences between them can yet again be mapped to the placement of the logical concept NOT. This also includes morpho-syntactic contronyms such as 'inflammable'. Wierzbicka (1996: 223) gives an elaborate description of the semantic molecule of fire, which if associated with some sort of object and particle, the NOT may also be clearly defined. Due to the nature of the explications available via the means of NSM, it is feasible to assume that this approach will be successful in relation to all contronyms.

Conclusions

This paper aimed to show the idiosyncrasies of natural language in the form of contronyms and their explication by means of sematic primes. The existence of the discussed sense opposition may prove to be problematic in the fields of not only theoretical linguistics but also natural language processing, artificial intelligence or machine translation. However, the phenomenon may be tackled by means of natural semantic metalanguage theory, which has proved to be successful on the grounds of cognitive linguistics.

When it comes to NSM theory, although one can observe "a trend towards increased systematisation and formalisation" (Goddard 2008: 2), it has not yet been formalized in a sufficient way to be viable in the field of computer science. However, if not this particular theory, then any other framework that would use primitive semantic carriers of meaning might prove successful. The work of the members of the so-called Moscow School of Semantics (see Apresjan 1992, 2000; Mel'čuk 1981) was based on similar principles and has already received attention in practical applications within the field of IT (see Fähndrich 2014 *et al.*). Nonetheless, the outline presented in this paper may become fruitful considering the claims of the major proponents of NSM theory,

which is "arguably approaching the standard expectations of a formalized metalanguage for natural language" (Goddard 2006: 544).

As a side note it can be said that contronyms help position natural language against formal systems such as formal languages, for example. In this perspective the natural language tends to have sense relations that are opposite to each other concealed behind one lexeme, and this radical polysemy can also be preserved in fully fledged sentences. This is definitely not a common behaviour among formal systems and, as mentioned at the beginning of the article, can be used as an example showing the stark differences between those systems and ethnic languages. However, one interesting to note fact is that from the perspective of NSM theory, the metalanguage embedded in all ethnic languages is in those terms similar to formal systems, as in this perspective the substance of *lingua mentalis* is based on semantic primes that have one meaning only. In other words, from this cognitive perspective it is possible to treat the human mind, at least on the level of semantics, as a coherent and non-contronymous system that is not based on antilogy.

To demonstrate the significance such entities like contronyms one may point to the project NELL (Never-Ending Language Learning system). This semantic machine learning system, developed by a research team at Carnegie Mellon University, was programmed to mimic the way humans absorb new data. It is supposed to learn by reading the content of the world wide web. The model of inferences is based upon: "(…) an initial ontology defining categories (e.g., Sport, Athlete) and binary relations (e.g., Athlete Plays Sport (x,y))" and "approximately a dozen labeled training examples for each category and relation (e.g., examples of Sport might include the noun phrases "baseball" and "soccer")" (Mitchell *et al.*). Grounded on such input the program is supposed to "(…) extract, or read information from the web to populate a growing structured knowledgebase, and (…) learn to perform this task better than on the previous day" (Carlson *et al.*). It becomes apparent that the inconsistencies caused by the phenomenon of contronymy within the natural language, if not dealt with correctly, must have an impact on projects like NELL. Nonetheless, the phenomenon is undoubtedly an interesting topic, showing the intricacies not only of language itself but also of human cognition, that calls for further investigation.

References

Apresjan, Jurij. 2000. *Systematic Lexicography*, trans. Kevin Windle, Oxford: Oxford University Press.

Carlson A., Betteridge J., Kisiel B., Settles B., Hruschka E.R., Mitchell T.M. 2010. Toward an Architecture for Never-Ending Language Learning. *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI'10)*,1306-1313.

Fähndrich, J., Ahrndta S., Albayraka S. 2014. Formal Language Decomposition into Semantic Primes. *Advances in Distributed Computing and Artificial Intelligence Journal* 3.8, 56-73.

Fillmore, Charles J. 1968. The case for case. In: Emmon. Bach & Robert T. Harms (eds.), *Universals in Linguistic theory*, New York: Holt Rinehart and Winston, 1-88.

Fillmore, Charles J. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, Conference on the Origin and Development of Language and Speech 280, 20-32.

Goddard, Cliff 2003. Thinking across Languages and Cultures: Six Dimensions of Variation. *Cognitive Linguistics* 14 (2-3), 109–140, DOI: https://doi.org/10.1515/cogl.2003.005.

Goddard, Cliff 2006. Natural Semantic Metalanguage. In: Keith Brown (ed.), *The Encyclopedia of Language and Linguistics*. 2nd edition. Amsterdam/Heidelberg: Elsevier.

Goddard, Cliff. 2007. Semantic molecules. In: Ilana Mushin & Mary Laughren (eds.), *Selected Papers of the 2006 Annual Meeting of the Australian Linguistic Society* at: http://espace.uq.edu.au/.

Goddard, Cliff. 2008. Natural Semantic Metalanguage: The State of the Art. In: Cliff Goddard (ed.), *Cross-Linguistic Semantics*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1-34.

Goddard, Cliff. 2010. The Natural Semantic Metalanguage Approach. In: Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, Oxford: Oxford University Press, 459-484.

Goddard, Cliff. 2012. Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology. In: Maria Koptjevskaja-Tamm & Martine Vanhove (eds.), *Linguistics. Special issue on "Lexical Typology"*, 50(3), 711-743, DOI: https://doi.org/10.1515/ling-2012-0022.

Goddard, Cliff, Wierzbicka, Anna. 2001. Język, kultura i znaczenie: semantyka międzykulturowa. In: Elżbieta Tabakowska (ed.), *Kognitywne podstawy języka i językoznawstwa*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych Universitas, 175-202.

Goldberg, Adele. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago/London: The University of Chicago Press.

Karaman, Burcu. I. 2008. On Contronymy. *International Journal of Lexicography* 21 (2), 173-192.

Klégr, Aleš 2013. The limits of polysemy: enantiosemy. *Linguistica Pragensia* 2, 7-23.

Lehrer, Adrienne. 1985. Markedness and Antonymy. *Linguistics* 21, 397-429.

Lutzeier, Peter R. 2007. *Wörterbuch des Gegensinns im Deutschen*. Band 1: A - G. Berlin/New York: Walter de Gruyter.

Matthews, Peter H. (ed.). 1997. *The Concise Oxford Dictionary of Linguistics*. Oxford/New York: Oxford University Press.

Mel'čuk, Igor A. 1981. Meaning-Text Models: A recent trend in Soviet linguistics. *Annual Review of Anthropology* 10, 27–62.

Minsky, Marvin. 1975. A Framework for Representing Knowledge. In: Patrick Henry Winston (ed.), *The Psychology of Computer Vision,* New York: Mvgraw-Hill, 211-277.

Mitchell, T., Cohen W., Hruschka E., Talukdar P., Betteridge J., Carlson A., Dalvi B., Gardner M., Kisiel B., Krishnamurthy J., Lao N., Mazaitis K., Mohamed T., Nakashole N., Platanios E., Ritter A., Samadi M., Settles B., Wang R., Wijaya D., Gupta A., Chen X., Saparov A., Greaves

M., Welling J. 2015. Never-Ending Learning. Proceedings of the Conference on Artificial Intelligence (AAAI). Association for the Advancement of Artificial Intelligence (www.aaai.org).

Spranger, M., Suchan J., Bhatt M. 2016. *Robust Natural Language Processing - Combining Reasoning, Cognitive Semantics and Construction Grammar for Spatial Language*. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), DOI: arXiv:1607.05968 (12 April 2018).

Wierzbicka, Anna. 1980. *Lingua Mentalis. The Semantics of Natural Language*. Sydney/ New York/ London/ Toronto/San Francisco: Academic Press Australia

Wierzbicka, Anna. 1996. *Semantics. Primes and Universals*, Oxford,New York: Oxford University Press.

Wolniewicz, Bogusław. 1980. Języki i kody. In: Adam Schaff (ed.), *Zagadnienia socjo- i psycholingwistyki*. Wrocław: Polska Akademia Nauk, Instytut Filozofii i Socjologii, Zakład Narodowy Imienia Ossolińskich, Wydawnictwo Polskiej Akademii Nauk.

RAOUL KARIMOV
Chelyabinsk State University

DOI: 10.15290/cr.2018.21.2.04

Combined Machine--Learning Approach to PoS-Tagging of Middle English Corpora

Abstract. This paper considers the problem of part-of-speech tagging in Middle English corpora (as well as historical corpora in general). Whereas PoS-tagging in general is now considered a solved problem for Modern English and is mainly achieved via hidden Markov models (HMM) and matrix-based word-to-vector conversions with every word in the dictionary being embedded into a single dimension, this approach relies on recurrent syntactic structures and context-free generative grammars and is therefore not applicable to older iterations of the English language due to irregular word order. As such, we believe that Middle English could be better handled by a morphographemic encoding and instance-based machine learning algorithms like SVM, random forests, kNN, etc. Using a moving-average method to generate multidimensional vectors giving a reliable numeric representation of character composition and sequences, we have achieved a precision and recall of 87.5% in classifying Middle English words by their part of speech while using a simplistic combined voting-based binary classifier. This result could be deemed satisfactory and encourages further research in the area.

Keywords: Instance-Based Learning, Corpus, Middle English, PoS-Tagging, Moving Average.

1. Introduction

Part of speech tagging is one of the central issues in the discipline known as natural language processing; it is quite frequently approached by means of the sequential-in-nature hidden Markov models that are basically designed to predict the probability of an event in a sequence given the previous events (Jurafsky 2008: 139). As Modern English follows a very strict word order, where, for instance, the definite article the and most other determiners are in over half of all cases followed by a noun, HMMs can be efficiently used to correctly classify words by their part of speech. Furthermore, PoS-tagging is assisted by finite-state transducers, which derive a given word's morphological properties by identifying grammatically-significant character sequences, or morphemes (Beesley and Karttunen 2004: 240). It is, however, a completely different story with Middle English, especially its earlier iterations (post-William the Conqueror's capture of Britain in 1066 and up to the issuance of Magna Carta Libertatum in 1215). The language, being characterized by less regular word order as well as rich morphology and very inconsistent orthography, where *scyl-*

de was the same was as *scilde* (shield), and *y-broht* was only slightly diachronically separated from *gebrouhte* (Pt. II of *bringan*, to bring), still presents a challenge for linguists working in the field of corpus linguistics and natural language processing, whereas not being extensively researched by machine learning community who prefer to allocate more resources, time, and scientific effort to modern languages.

We have, unfortunately, so far found no previous research what would be dedicated specifically either the lemmatization or the PoS-tagging of corpora in Middle English. Neural networks have been made for Slavic and other morphologically complex languages (Jędrzejowicz and Strychowski 2005: 200; Malouf 2016: 123), but utilized better-codified and larger-in-volume language data than we could ever afford in our Middle English effort. Nevertheless, PoS-tagging and other functions provided by NLP applications could be of great use for philologists studying language history who are currently restricted to corpora with limited annotation and as such have to perform annotation manually, which is both time-consuming and labor-intensive, thus reducing the overall efficiency of scientific work. With that in mind, we decided to find a way to automate the process of PoS-tagging by applying existing machine learning methodology.

For this study, the following was hypothesized: there should exist a simple instance-based machine learning method that would enable efficient PoS-classification of orthographically volatile Middle English words while trained on a small set of data. We believed that support vector machines (SVM), random forest models (RFM), k nearest neighbors (kNN), and Bayesian algorithms could all be used for such learning. The hypothesis is to be verified by means of 10-fold cross-validation, whereas the difference in results returned by various algorithms is to be t-tested by Student's method.

2. Theory and methodology

2.1. Research data

This study derives data for analysis primarily from the Helsinki Corpus of English Texts (Web 1), which contains about 450 texts of philosophical, literary, belletristic, epistolary, scientific, and religious nature, a total of 1.5 million words. The corpus is freely available via the Oxford Text Archives and is subdivided into multiple smaller divisions based on the diachronic classification of its texts, which in its turn is dual, as it considers both the creation of the original text and the creation of its manuscript that the corpus includes. Preliminary analysis of the corpus data and the preparation of training and test samples were done by consulting Mayhew and Skeat's *A Concise Dictionary of Middle English* (Mayhew and Skeat 1888). For the goal of this research paper, we limited ourselves to one of the texts from the corpus: Vespasian Homilies, ca. 1167, which partially reduced the overall orthographic and grammatical inconsistency that could be observed across the dialects of that time.

For the initial machine learning effort, we would like to perform simple binary classification to see what results are reasonably achievable on a smaller training set. For this purpose, we made a

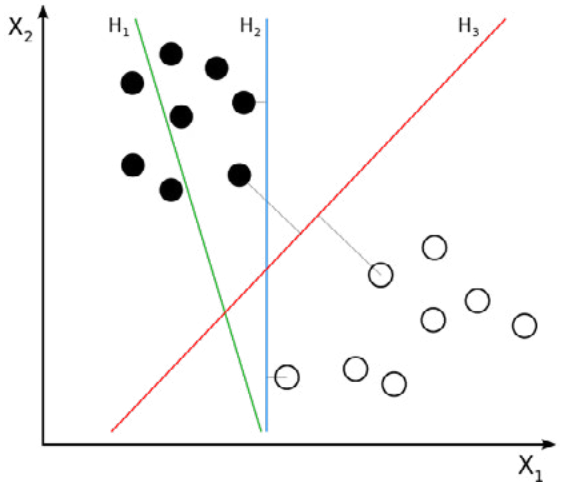
small 200-word set containing 110 verbs and 90 adjectives, all extracted from Vespasian Homilies as they are contained in the text, i.e. without lemmatization, spelling normalization, or any other pre-training rectification, which we believed would be inappropriate given that the ultimate goal of the research effort was (and is) to develop a technique that could be applied to raw corpus data. The set is made deliberately small so as to better optimize the methodology for training on limited data.

2.2. Algorithms Under Investigation

In this on-going research effort, we are investigating the capacities of several known instance machine-learning algorithms when applied to small Middle English vocabulary samples (for space considerations and for the sake of simplicity, we are not providing any detailed mathematical descriptions of those algorithms, see sources cited): Support Vector Machines, **K Nearest Neighbors, Random Forest Models, and** Multilayer Perceptron.

A support vector machine, or SVM, is in its most basic form (a binary linear machine) an algorithm that, given a set of vectors in *n*-dimensional vector space, finds a hyperplane that maximizes the margin between itself and the vectors on either side of it. Figure 1 below gives a good illustration of how it works:

Figure 1. Separation of Vectors

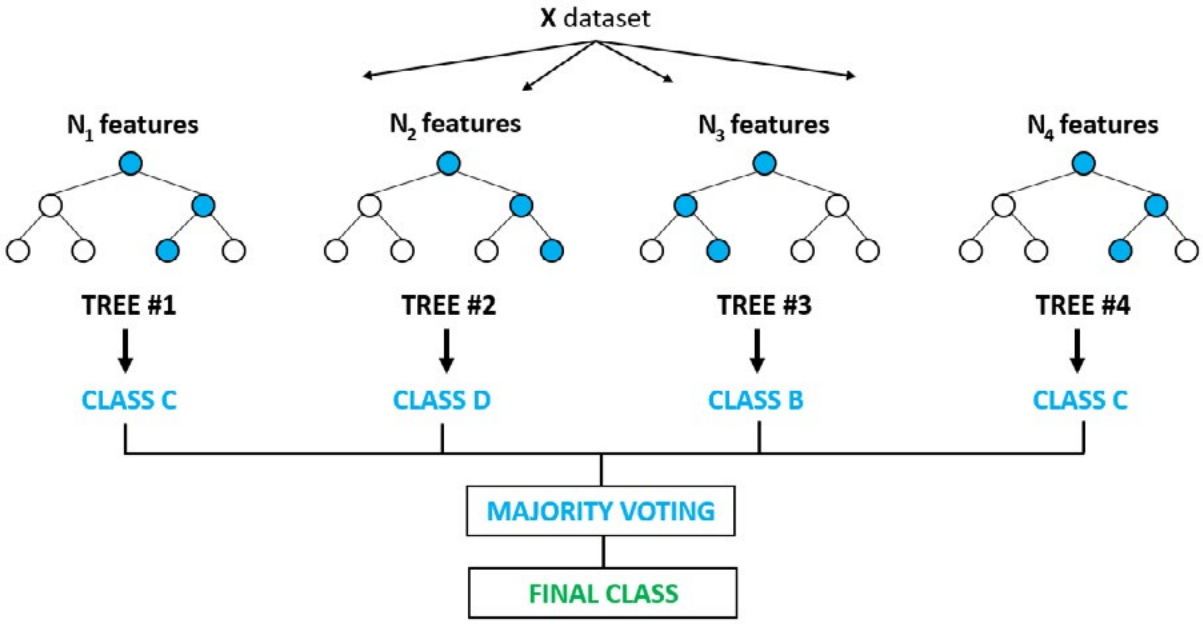


In the figure above, black and white dots represent linearly-separable vectors of different classes; the green line does not separate the classes; the blue one does with a minimum margin; the red one does with a maximum margin and is therefore deemed the best hyperplane. In case of linearly separable data, the maximum-margin hyperplane lies between two other planes that border the separated vector sets; vectors from the data set that belong to such other planes are called support vectors, hence the name of the method; for a detailed mathematical description, see (Christianini and Shawe-Taylor 2000: 94).

Random forest models, or RFM, generate decision trees, each of which is applied to classify a random subset of the training sample using a random partial set of attributes, or features (neither

the sample, nor the attributes are used entirely). When classifying a new instance, RFM gives preference to the class which the majority of such randomly-generated trees has voted for, i.e. performs simple majority voting (Breiman 2001: 2). Of the randomly-reduced set of attributes, the best one that is used for generation of tree nodes is often chosen on the basis of Gini impurity. Below is an illustration of how it works:

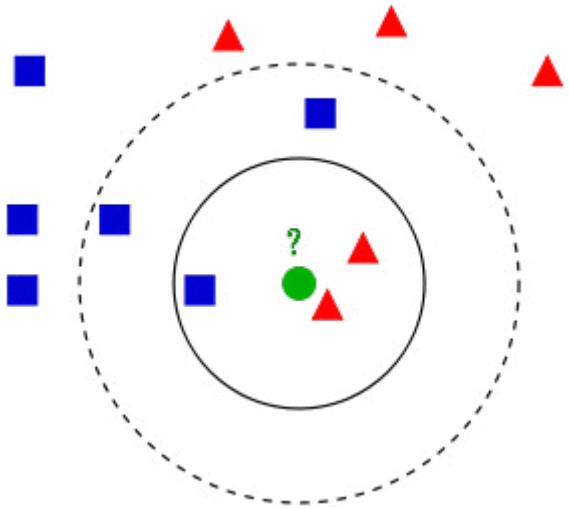
Figure 2. Random Forest Model



As can be seen in Figure 2, each decision tree is fed a part of the data set, drawn at random, and uses a random subset of features to perform the classification of an unknown-class instance. In tasks other than classification, e.g. regression, averaging can be used instead of majority voting. The random forest algorithm scales well, is efficient for any number of attributes or classes, is optimal for large and small data alike, and is good for both continuous and discreet attributes; while being computationally intensive is often cited as its main drawback, it does not seem to be problematic given our purpose to use deliberately small training sets.

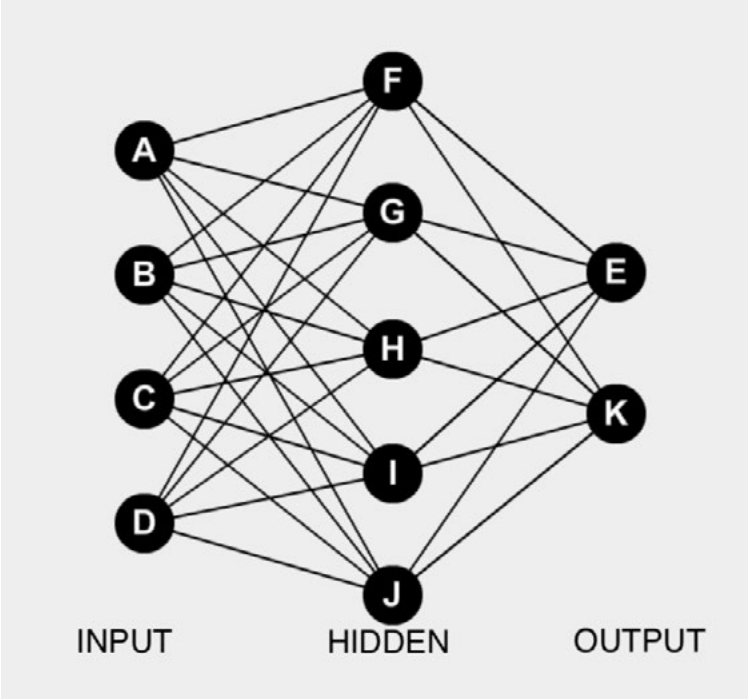
K nearest neighbors is one of the simplest yet very efficient instance-based learning algorithms and classifiers that uses only basic majority voting (Aha 1991: 38). A new instance is vectorized and then placed in an n-dimensional space of correctly classified vectors (e.g. the training-sample vectors), whereby the algorithm calculates the Euclidean or Manhattan distance from this new vector to the known-class vectors to identify a number of nearest neighbors. The number is usually odd and is referred to as k, hence the name of the algorithm. Whichever class has the most vectors among those k nearest neighbors, that class is assigned to the new vector as well. In Figure 3 below, the green vector in the center would be classified as a red triangle if $k = 3$, or as a blue square if $k = 5$.

Figure 3. K Nearest Neighbors



Multilayer perceptron, or MP, is the most complex algorithm of the above, and is a feedforward artificial neural network (Teijiro et al. 2012: 758); unlike SVMs, it can be used on data that is not linearly separable. It builds a network consisting of three or more layers (input, output, and at least one hidden layer) containing nodes with weights adjusted by a sigmoid function. Weight adjustment is done by the backpropagation of error. MLPs are successfully used in scientifically complex prediction problems like Seyed (2015: 62).

Figure 4. Schematic representation of an MLP with three separately-weighted hidden ode layers



Middle English, despite being rather inconsistent both grammatically and orthographically, did have regular morphs that are still referred by historical linguistics as the primary categorial markers. Hypothetically, if we were able to generate word-vectors such that similar character sequences occurring in similar intra-lexeme positions would produce closely-positioned vectors, then a vector-based machine learning algorithm such as SVM or an RFM should be able to correctly link together words that have similar graphical clusters at the beginning (the prefix) and/or at the end (the suffix), which in many (though not all) cases would suffice for part-of-speech classification.

2.3. Methodology

As has already been mentioned above, the method to use had to focus on the recurring sequences of symbols observed in words and signifying its PoS-category. As such, we had to find a simplistic yet efficient method that would enable us to represent words in a vector form that would reliably shaped by both the character composition of, and character-specific position in, a given word. Takala cites several methods of vector-word embedding: regular embedding (one dimension per word for n top-ranking tokens contained in the dictionary, with the rest being assigned to a single dimension associated with rare words), stem+ending embedding (stem vectors concatenated with ending vectors), and moving-average method that uses relatively small dimensionality to collect information from all parts of a word (Takala 2016: 179). In our pursuit of making such representations that would not require any pre-specified rules while also not being calculation-intensive, we decided to choose the latter.

The moving-average representation is essentially a vector of n dimensions, where n = number of characters in the alphabet, with each dimension being assigned to a single character. A word representation $w = (w_a, w_b, \ldots w_z)^T$:

$$W_\alpha = \sum \frac{(1-\alpha)^c \alpha}{z}, \quad (1)$$

where *c* is the character index (1 for the first symbol in the word, 2 for the second one, etc.), α is a hyper-parameter to control the decay, and Z is a normalizer proportional to the word length (which we decided to be the word length itself, i.e. 4 for word). Thus, each word-vector contains a weighted sum in each dimension representing any character that found in the word, and 0 in the rest of dimensions. The operation is repeated backwards, and the new vector is concatenated to the previous one so that *word* is represented as *word* ⁀ *drow*. Takala also suggests concatenating a third vector which only contains character counts; for now, we decided not to use that option.

Before the experiment was conducted, we had done a limited normalization of spelling: both thorn and eth had been replaced with the cluster *th*, whereas ash, æ, had been replaced with *ae*, and yogh, ȝ had been replaced with *g*. We also removed diacritics and decapitalized all the words in both of our sets. Thus, we came to an alphabet of 26 characters, which resulted in word-vectors in a 52-dimensional space over the field of real numbers (26x2). Two sets were then made, each containing 100 instances of 52 numeric attributes and one binary attribute POS {VERB, ADJ} to test the ability of support-vector machines to effectively differentiate these two parts of speech

based on instances extracted from a Middle English Text. All machine learning algorithms were run in the Weka data-mining environment (Frank and Witten 2016: 365).

3. Experimentation and discussion

As mentioned above, training and verification by 10-fold cross-validation were performed on a small 200-word sample containing 110 verbs and 90 adjectives from a single Middle English texts. All the four models discussed in Section 2.2 were first trained and verified separately; below are the confusion matrices for all models.

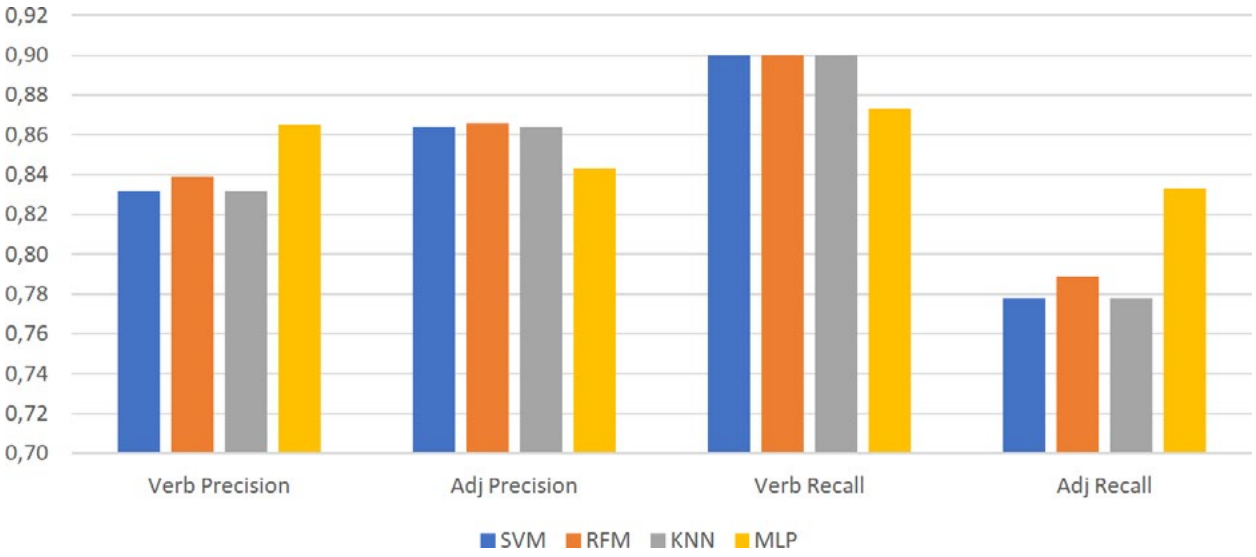
Table 1. Confusion matrices for each model

MLP			**kNN**			**SVM**			**RFM**		
verb	adj	cas	verb	adj	cas	verb	adj	cas	verb	adj	cas
96	14	verb	99	11	verb	99	11	verb	99	11	verb
15	75	adj	20	70	adj	20	70	adj	19	71	adj

Note: cas means "classified as". Values in bold are the best values achieved across algorithms.

To better illustrate how precise each algorithm was in separating verbs from adjectives, we plotted a precision-and-recall histogram for each of them, see Figure 5 below.

Figure 5. Class-specific precision and recall values for different algorithms



As can be seen from the graph, adjectives are universally classified more precisely than verbs but show considerably lower recall. Interestingly, kNN and SVM return identical results for both classes; verb recall and adjective precision have the same or nearly same values across algorithms except MLP, which is superior to others in terms of verb classification precision while somewhat worse in terms of verb recall. It is also the only algorithm to show high recall for ad-

jectives at the cost of somewhat lower precision. Table 2 provides detailed classification results for each algorithm.

Table 2. Detailed accuracy by class for each model

TP Rate	FP Rate	Precision	Recall	F-Msr	MCC	ROC Area	PRC Area	Class
Support Vector Machines								
0.900	0.222	0.832	0.900	0.865	0.687	0.839	0.804	VERB
0.778	0.100	0.864	0.778	0.819	0.687	0.839	0.772	ADJ
0.845	0.167	0.846	0.845	0.844	0.687	0.839	0.790	Weighted Average
Multilayer Perceptron								
0.873	0.167	0.865	0.873	0.869	0.707	0.909	0.919	VERB
0.833	0.127	0.843	0.833	0.838	0.707	0.909	0.872	ADJ
0.855	0.149	0.855	0.855	0.855	0.707	0.909	0.898	Weighted Average
Random Forest								
0.900	0.211	0.839	0.900	0.868	0.697	0.940	0.951	VERB
0.789	0.100	0.866	0.789	0.826	0.697	0.940	0.923	ADJ
0.850	0.161	0.851	0.850	0.849	0.697	0.940	0.938	Weighted Average
K Nearest Neighbors								
0.900	0.222	0.832	0.900	0.865	0.687	0.839	0.810	VERB
0.778	0.100	0.864	0.778	0.819	0.687	0.839	0.767	ADJ
0.845	0.167	0.846	0.845	0.844	0.687	0.839	0.790	Weighted Average

In this table, values in bold are the best precision, recall, and ROC values achieved by means of all the four algorithms tested. Note that the power of algorithms is highly contextual: MLP is better at handling adjectives, whereas RFM returns higher ROC for any class, and SVM/kNN deal with verbs more efficiently. This suggests that different algorithms might perform better or worse depending on the class, the metric, or other factors, which is why we propose making a combined

algorithm. To that end, we decided to combine all the four algorithms to make a majority-voting meta-classifier, whose results are shown below.

Table 3. Detailed accuracy by class for the majority-voting classifier

TP Rate	FP Rate	Precision	Recall	F-Msr	MCC	ROC Area	PRC Area	Class
0.909	0.167	0.870	0.909	0.889	0.747	0.871	0.841	VERB
0.833	0.091	0.882	0.833	0.857	0.747	0.871	0.810	ADJ
0.875	0.133	0.875	0.875	0.875	0.747	0.871	0.827	Weighted Average

Apparently, using a combined classifier did improve both precision and recall for both classes, which suggests it is a recommendable approach for further machine-learning efforts in this research. A weighted-average precision and recall of 0.875, we believe, indicates that the combined model showcases a sufficient capability of predicting the part of speech of a given word when trained on 52-dimensional word-vectors generated by the moving-average method. However, a few problems should be highlighted.

First it would be useful to note that verbs generally demonstrate better results than adjectives with all the algorithms, which we think is due to the sampling method: as we did not lemmatize or otherwise normalize the form of words we tested the approach on, some adjectives in both sets were given in the superlative form. Given the absence of superlative rhoticism (Ilyish 1968: 104) in Middle English, such adjectives would bear the suffix *-st-*, which coincided with the verbal 2SG suffix; as second-person verbs were abundant in the sample due to the nature of Vespasian Homilies, where the author directly addresses the reader and uses the 2SG personal pronoun þu, this could have resulted in a consistent association of the character string *-st-* with verbs, which affected superlative adjectives.

Second, it should be borne in mind that the experiment was oversimplified and reduced to two parts of speech, one of which (the verb) is known to be the most morphologically complex and rich in most languages of the world, thus boasting better and more indicative character-string markers. On the other hand, ME nouns and adjectives did share many of their case-specific suffixes, which would probably result in multiple confusions of these two parts of speech should both be included in the experiment. This might mean that the included algorithms might not make a sufficient PoS-tagging tool despite the morphological richness of Old and Middle English, necessitating the implementation of non-character-based means, e.g. hidden Markov models.

Another question that is yet to be answered is whether multi-class machines would be as efficient as binary ones. SVMs are primarily intended for binary classifications and utilize a breakdown approach when it comes to more than two classes; MLP might behave differently on a larger sample. Further experiments are needed to find out if they could be efficient in our case should

we try to train those models and/or the combined classifier to recognize all parts of speech. This is complicated by cross-PoS homography that could sometimes be observed in historical English texts, where words of different parts of speech could have the same or near-same graphical representation. This reiterates the need for including non-morphology-based techniques such as HMMs, which, however, need a different vector representation.

Finally, we have to mention the grammatical ambiguity that has existed since the appearance of analytical perfective structures in the English language, where the second participle of a verb could be used as both a non-finite component of a verbal compound, or as an attribute, in which case it would fulfill the role of an ordinary adjective. This is essentially a problem of grammatical classification rather than machine learning, yet it could adversely affect the latter.

4. Conclusions

This paper analyzes multiple classifiers and a combination thereof that use multidimensional word-vectors generated by means of a moving-average formula applied to every word in a set in direct and reverse order to create a vector reflecting both the character composition of, and the weighted character-specific position in, a given word. All the models are cross-validated on a small verb-and-adjective set sampled from a Middle English corpus. Despite somewhat worse results for adjectives, a combined voting-based classifier compensates for it by taking into account the MLP output, where VERB/ADJ precision and recall values are balanced. The results obtained encourage further research and experimentation in this area; however, they probably indicate the need for complementing the approach with another model that does not rely on character strings, e.g. HMM. This will be the basis for our future efforts in the pursuit of creating a model that is able to correctly classify words by their part of speech or other morphosyntactic properties, in Middle English and other historical languages that share the same properties: morphological richness and orthographic inconsistency. The current findings, however, are deemed satisfactory given the small sample size.

References

Aha, David W., Kibler, Dennis, Albert, Marc K. 1991. Instance-based learning algorithms. *Machine Learning* 6-1, 37-66.

Beesley, Kenneth R., Karttunen, Lauri. 2004. Finite-State Morphology. *Journal of Computational Linguistics* 30-2, 237-249.

Breiman, Leo. 2001. Random Forests. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf (19 April, 2018).

Christianini, Nello, Shawe-Taylor, John. 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press.

Frank, Eibe, Witten, Ian H. 2016. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: Morgan Kaufmann.

Ilyish, Boris A. 1968. *History of the English Language*. Moscow: Vysshaya Shkola.

Jędrzejowicz, Piotr, Strychowski, Jakub A. 2005. Neural Network Based Morphological Analyser of the Natural Language. *Intelligent Information Processing and Web Mining. Advances in Soft Computing* 31, 199–208.

Jurafsky, Dan, Martin, James H. 2008. *Speech and Language Processing.* New Jersey: Prentice Hall.

Malouf, Robert. 2016. Generating morphological paradigms with a recurrent neural network. *San Diego Linguistic Papers* 6, 122–129.

Mayhew, Anthony L, Skeat, Walter.1888. *A Concise Dictionary of Middle English From A.D. 1150 to 1580.* Oxford: Clarendon Press.

Seyed, Hamid H., Mahdi, Samanipour. 2015. Prediction of Final Concentrate Grade Using Artificial Neural Networks from Gol-E-Gohar Iron Ore Plant. *American Journal of Mining and Metallurgy* 3-3, 58-62.

Takala, Pyry. 2016. Word Embeddings for Morphologically Rich Languages. *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 177-182.

Teijiro, Isokawa, Naruhiko, Nishimura, Nobuyuki, Matsui. 2012. Quaternionic Multilayer Perceptron with Local Analyticity. *Information* 3, 756-770.

Web 1 – Helsinki Corpus of English Texts.www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus (4 April, 2018).

EZEKIEL OPEYEMI OLAJIMBITI
Elizade University, Nigeria

DOI: 10.15290/cr.2018.21.2.05

Discourse Pattern, Contexts and Pragmatic Strategies of Selected Fraud Spam

Abstract. The thrust of this paper is the pragmatic investigation of fraud spam, the unwanted emails containing the strategic use of language with the intention to swindle money from the recipients. Sixty (60) English medium email samples were collected from the author of the present paper's email spam between July 2017 and February 2018 in Nigeria. These were analysed using Halliday and Hasan's Generic Structure Potential and an aspect of Fetzer's cognitive context model. The study identified six discourse patterns: salutation, discourse initiation, enticing information, mild conscription into business, request and subscription; orienting to contexts of business and religion; manifesting pragmatic strategies of adversatives, evocation of business idea, evocation of religious affinity and evocation of messianic figure. The study, therefore, concludes that cyber-fraudsters deploy similarly familiar patterns and contexts evincing strategic persuasive language to defraud their prospective victims. Significantly, the study complements existing literature on fraud discourse in linguistic scholarship.

Keywords: Fraud spam, generic structure potential, mild conscription, messianic figure and cyber-fraudster.

1. Introduction

The ubiquitous nature of email fraud, the unwanted emails containing the strategic use of language with the intention to swindle money from the recipients, is a continuous social concern worthy of scholarly investigation. Studies have shown that millions of people, the world over, have become and still fall victims of this form of cybercrime because the fraudsters have not desisted from sending such emails. I was almost a victim a few years ago when I ignorantly responded to one of the emails in my spam account; at the verge of taking a loan in an attempt to raise the money requested by the fraudsters, but for the intervention of one of my acquaintances who divulged to me the secrete of cyber-fraudsters. The foregoing motivated this study, the pragmatic investigation of fraud spam, with a view to determining the discourse patterns, contexts and pragmatic strategies that are often evinced in the criminally oriented and fraudulently characterised emails allegedly sent to recipients to swindle money from them.

Hua, Abdollahi-Gullani and Zi (2017) aver that the issue of the victim's vulnerability has continued to be on the increase and the reasons behind it certainly deserve further linguistic and

Metalinguistic scrutiny. Studies in this direction have not been widespread due to the search for appropriate methodologies. Significantly, the existing studies have mostly covered linguistic discourse analytical approach (Chiluwa 2009, 2013; Barron 2006; Orasan and Krishnamithy 2002; Heyd 2008). The present study, therefore, complements these studies by its adoption of a discourse-pragmatic approach to email fraud (cyber fraud) in linguistic scholarship in Nigeria.

Email (electronic mail) is one of the most frequently used services in the net. Runkehl, Siever and Schlobinsk (1998) claim that email is quicker than ordinary post service, cheap, and can be sent to a lot of addresses simultaneously. This seemingly friendly phenomenon has been greatly abused for fraud purpose. Email fraud includes fake lottery winning announcement or fake business invitation, money transfer, investment opportunity, dormant account claim invitation, money inheritance information (Chiluwa 2013:1). According to him "in the last ten years, email as a form of computer-mediated communication (CMC) has increasingly become the main medium for perpetrating digital deceptions, email fraud or digital lies.

Fraud, according to Idowu (2009: 629), is the deliberate falsification, camouflage, or exclusion of the truth for dishonesty/ stage management to the financial damage of an individual or an organization. It is the dishonesty or an act of cheating someone or business to give up possession of some lawful right (Polick 2006). In their view, Fadipe-Joseph and Titiloye (2012: 215) refer to fraud as any actions by which one person intends to gain a deceitful advantage over another.

2. Related studies

The continuous rise of cyber fraud cases demands serious preventive actions that would emanate from scholarly investigations. Chiluwa (2009) investigates the pragmatics of hoax email business proposals; where he applies speech acts theory to the study of discourse strategies and functions of hoax email. He classifies this as "419 emails", being the Nigerian term form all forms of online fraudulent acts. His submission is that such practices have become a regular part of our internet social life since economic hardship being witnessed by the world today can force people to criminal activities.

The attention of researchers has continued to be drawn to the strategies deployed by these fraudsters to swindle money from their intending victims. For instance, Behman, Azabdaftari and Hosseini (2011) identify persuasive strategies of personalization, presupposition and lexical choices as frequently being used by fraudsters in financial fraud spam emails. In a similar vein, Hiss (2015) in his study on fraud and fairy tales: storytelling and linguistic indexicals in scam emails, opines that most of us receive numerous spam emails, text that in one way or another try to convince us to engage in transactions of enormous sums of money, promising unbelievable benefits. He argues further that in the fraudsters' attempts to get the recipients involved combining cultural indexicals, interactional roles and narrative strategies. Recently, Hua, Abdollahi-Gullani and Zi (2017), explore linguistic deception of Chinese cyber fraudsters. They argue that cybercrimes are on the increase in China, where fraudsters manipulate language to deceive users into revealing their bank account or depositing sums in the cheats' account. The study adopted speech acts and politeness theories to conversations of 50 interlocutors who had already chatted with different on-

line cheats in China. It submits that fraudsters manipulate language to present untruth as truth using online deceptions.

Proffering solution, Yoon et al. (2010:12) propose a hybrid spam filtering framework for email communication using a combination of contact-based filtering and challenge-response. The study performs some preliminary experiments to investigate whether plagiarism direction tools could be used as a filter form of spam emails. In a related view, Leap (2007:63-64) points out that email fraud may have either one of two objectives, first, fraudulent emails techniques may be used to cover up the misappropriation or misapplication of funds. Secondly, fraudulent financial statements may be issued by email writers to mislead the mail recipients. An overview of the literature, however, shows that the myriad of techniques can be broken down into four broad categories the first two are revenue based schemes and expense based schemes the two revenues are aimed at anticipating a firm current profitability as reported on the income statement. The third and fourth categories are asset-based schemes and liability-based schemes. Scholars like Young (2006) and Dechow et al., (2011) have, however, suggested that the primary motivation for email fraud is only a small subject of cases. Those fraudsters also use well-known business entities or organisations and charity instinct to deceive the recipients.

3. Theoretical perspectives: Generic Structure potential and cognitive context

3.1. Generic Structure potential

Generic Structure Potential directly derives from Contextual Configuration (CC). According to Odebunmi (2007: 88), "Generic Structure Potential (GSP) emerges from Halliday and Hasan's (1989) concept of contextual configuration, which has been added to the earlier Hallidayan context of situation dimensions". These functional linguists argue further that field is the social action or what is going on, and language is playing a significant role. It determines the register used. The tenor of discourse is the cluster of a meaningful relationship between relationships the participants in acts communication. CC gave prominence to the immediate constituents of a text, with no serious consideration to all the potentials of a genre. Therefore, GSP was developed to take into cognizance, all the possibilities that can occur in a text- obligatory, optional and recursive elements in possible orderings. It also refers to the staging of the genre at its attending sequencing and formalisation within the cultural experience. So it takes into cognizance all the features of texts possible by locating such within a specific genre if its structure is compatible with one of the possibilities specified by the GSP. The notations and their meanings are as follows: the dot (.) indicates more than one option in sequence, the round brackets ( ) represent optionality of enclosed elements, the square brackets [ ] show restraints of sequence, the braces with curved arrow { } indicate that the degree of interaction for elements in braces is equal, the caret sign ^ shows sequence. The choice of GSP to this study is significant in unpacking the discourse patterns of the

selected fraud emails. The structure of the theory allows for the sequences that characterise the content of email messages.

3.2. Cognitive context

Fetzer's contribution to cognitive context is adopted in this study. Cognitive context is a structured multilayered construct which is indispensable for language processing and inferencing. It is required for a cognitively based outlook on communication as it contains assumptions about mutual cognitive environments. According to Fetzer (2007: 12), "the nature of the connectedness between its constative layers and subsystems is meta-communicative and meta-systemic". Sperber and Wilson (1986: 95) stress that cognitive context refers to a set of premises, namely, true or possibly true mental representations. Deductive reasoning from the foregoing is that mental construct or representations on practical reasoning on language use are central to cognitive context. Fetzer identifies elements of this context as mental representations, propositions and contextual assumptions which may vary in strength and factual assumptions. Features of this context are deployed in accounting for the contexts the assumed cyber-fraudsters orient to by cognitively appealing to the psyche of the email recipients to defraud them.

4. Methodology

Data for the study consist of sixty (60) English medium email samples collected from the author of the present paper's email spam between July 2017 and February 2018 in Nigeria. These emails were retrieved from the spam to show that they were sent from the unknown individuals. Their contents reflect that they are crime-related manifesting money transfer, lottery, business and charity engagements. These were analysed using Halliday and Hassan's Generic Structure Potential to unpack the discourse patterns of the data and an aspect of Fetzer's cognitive context model to track the context the cyber-fraudsters orient to in defrauding the email recipients and pragmatic strategies often deployed for such intention.

5. Data analysis and findings

The analysis is structured into discourse patterns, contexts and pragmatic strategies. These are analysed in turn.

5.1. Discourse patterns of fraud emails

Six discourse patterns characterise the selected fraud emails reflecting the language use of the cyber-fraudsters, namely, Salutation (ST), Discourse Initiation (DI), Enticing Information (EI), Mild Conscription into Business (MCB), Request (RQ) and Subscription (SB). These are catalogued thus:

[ST] ^ [DI] ^ [EI] ^ [MCB] ^ {RQ} ^ [SB]

The catalogue shows that all the elements are obligatory as one stage sequences to another in fixed positions representing the discourse patterns of email fraud. These are analysed in turn.

5.1.1. Salutation (ST)

Salutation describes phatic communication in routine greeting addressed to the email recipients. Four forms of salutation characterise selected data are polite concentration seeking method, pious greeting, polite routine observance and social tie marker. The polite concentration seeking method relates to a form of greeting addressed to the email recipients to get their attention by using honorific and polite terms like 'sir'. An example of this is "Attention Sir". Pious greeting explains a Christian way of showing affinity with a person. Salutation like "Dearest in Christ" is intended to appeal to the religious minds of the recipients. Polite routine observance connects to the ritual-like function of language but in this case, the cyber-fraudsters affiliate to social tie and romance markers like "friend", "dear", "love" to dress the positive face want of the recipients to defraud them.

5.1.2. Discourse initiation (DI)

This elucidates how the email writers introduce themselves to their recipients, entailing self-presentation reflecting identity terms of name, sex and one's occupation. Noticeable in the selected data are the uses of pronouns: personal and possessive pronouns and fictitious personal details of the email writers. These are exemplified below.

Example 1

'I am Dr Takashi Shimada, Japanese origin

'I am Mrs Mary Martins, an ageing widow'

'My wife and I won the euro millions lottery of J53 million'

'My name is Alice Joe from UK'

The foregoing indicates first person and possessive pronouns through which email writers get themselves introduced to the recipients. Such pronouns are immediately followed by fictitious names of the cyber-fraudsters, who also disclose their social identity reflecting social status and national identity. These implicate self-presentation through identity acknowledgement to authenticate the emails but such strategy constructs pragmatic nuances of deception.

5.1.3. Enticing information (EI)

Enticing information projects persuasive narratives by the email writers with the intention to convince the email recipients to be interested in the conversation and thereby swindle money from them. Such narratives revolve around business, charity, health, friendship money transfer and ATM card. The following samples exemplify this further.

Example 2

'I have some funds I inherited from my loving husband Mr Martins J martins the sum us &3.800.000.00. (Mail 2)

Example 3

'I am … a woman who was diagnosed for cancer about 4 years ago I have decided to donate my fund ($ 2,000,000.00) (Mail 6)

Example 4

We have voluntarily decided to donate $2,000,000.00 (Two Million Dollars) to 5 individuals randomly as part of our own charity project.

Example 5

Friendship is a gift. It is a blessing that only the fortunate have. I am lucky to have a friend like you…

Example 6

Right now we have finally succeeded in getting your ATM CARD worth of $1.5 million out of delivery your ATM CARD with the help of Adams Mole Attorney General of Federal High Court of Justice Benin Republic…

Example 7

I am here to search for a business partner or friend who will help me to invest my fund in his company or country.

The foregoing evinces different narratives ranging from business, charity, ATM card, friendship related matters. These, though, look ordinarily convincing to engage the email recipients to initiate the process of defrauding them. Expectedly, this stage of the email arouses and activates the interest of the recipients as it creates psychological prominence in their thinking faculties.

5.1.4. Mild conscription into business (MCB)

The mild conscription into business stage describes amiably gentle encouragement and provocation of an invitation to lure an individual into business. Business in this context extends beyond buying and selling to any engagement of human beings. Therefore, MCB characteristically explains an allurement, enticement, or attraction of a written request for someone's presence or participation. Its manifestation in the selected data reflects assistance often solicited from email recipients in terms of fund management, business establishment and friendship engagement.

Example 8

We need your assistance by representing my company in Nigeria

Example 9

I need a very honest and God fearing person that can use these funds for God's work and 15% out of the total funds will be for your compensation…

Example 10

Beloved let us join hands together to help our fellow brothers and sisters who are poor, sick and homeless so that blessings will be ours while glory goes to the Lord our creator.

Example 11

Although, I just want us to become friends maybe something more if you wish, and I want you to write me back for more discussion with you as soon as you receive this email.

This trend is obviously noticeable in the sampled data. The email recipients are mildly lured into company representation, fund management and friendship as traps to engage them in further discussions that would lead to syphoning money from them. Importantly, these narratives are intended to active attitudinal disposition in the recipients.

5.1.5. Request (RQ)

This obligatory element captures the act of asking or employing email recipients to do something by responding to emails received to indicate interest or give more information on personal details with the intention to facilitate further discussions. This has been classified into two, namely, information request and inherent request. Information request describes the method the cyber-fraudsters deploy to know the identity and other personal details of the email recipients. Examples of such include: "reply us with your resume/CV", "reply back the message he will respond to you immediately", "please reply me back". These convey strategic means of getting the email recipients engaged to aid their dubious intention. Interest request in the same vein shows how fraudsters want the email recipients to be committed by positively responding to the enticing information already given. This is captured in the following: "I will give you more details if you show more interest", "Please kindly get back to me as soon as possible if you are interested", "Please if you would be able to use the funds for the charity works, kindly let me know immediately". These are politely constructed to appeal to the positive face want of the email recipients, especially with the frequent use of "please".

5.1.6. Subscription (SB)

Subscription signals the closure of an email as a similitude of formal correspondence. Usually, it entails closing remarks and marker of authorship of the emails. Three forms of subscriptions identified in this study are formal closure ("Yours sincerely, Dr Takashi Shimada, Executive Director"), pious closure ("Your sister in the Lord Mrs Martins"), informal closure ("Best regards, Alice"). One noticeable trend is that the salutation and subscription follow similar narrative features.

5.2. Contexts and pragmatic strategies

Two contexts namely, business and religion, configure the selected data. These heavily manifest in all the stages presented in the discourse pattern section. Of significance is the fact that the cyber-fraudsters orient the recipients of the emails to these contexts to trigger a positive response from them. These are analysed along with the following pragmatic strategies: adversatives, evocation of business idea, evocation of religious affinity and evocation of messianic figure.

5.2.1. Context of business

This context indexes linguistic features pointing to commercial engagement involving the production, buying and selling of goods and services and other monetary matters. Fraud emails writers often orient their intending victims to fictitious economic activities like investments, company representation, lottery related matters and so on with the intention to activate psychological relevance in them. This is exemplified in the following:

Example 12

I am the executive director American Devices and Diagnostics Manufacturers Association (AMDD), we specialize in the production of interventional cardiology product and other devices… we need your assistance by representing the company in Nigeria.

Example 13

I am here to search for a business partner or friend who will help me to invest my fund in his company or country.

Example 14

..will you like to come and work and live in the USA?

In Example 12, the following co-texts lexicalise business orientations: "Diagnostics", "Manufacturers", "production", "cardiology product", "representing" and "company" which also capture company representation. The co-tests in Example 13, are "business", "invest", "fund" and "company" all pointing to the idea of investment. These are put together to appeal to the cognitive minds of the recipients who could relate easily to the mental representations of a business idea. Mental representation is a key term in the cognitive context that explains a presentation to the mind in the form of an idea or image that can be perceived. These cyber-fraudsters conjure mental representation of huge monetary opportunity to their recipients whom they have seen indirectly as lower socioeconomic status people.

Two pragmatic strategies associated with this context are evocation of business idea and adversative. The former relates to creating an image of the fantastic business idea with huge monetary opportunity. Significantly, the strategy configures mental representations of an ideal business world, while the recipients are expected to infer this by contextual assumptions from the contents of the emails. Contextual assumptions as used in cognitive context framework refer to

meaning inferred from the stated utterances. The latter entails information packaging act for a business purpose which in most cases is to win more "customers" in a competitive business world. Ong (1981: 51) argues that "'adversatives' is universal, but 'conspicuous or expressed adversatives is a larger element in the business world". This is illustrated in this example: "I would like you to handle the contract of supplying a product to my company". In this case, the fraudsters deploy this strategy to deceive their intending victims who may want to assume that such adverts are real.

5.2.2 Context of religion

The context of religion relates to linguistic configuration characterised by a system of belief in a being and the activities that are connected with this belief system. In the context of this study, the Christian religion is portrayed. This is evidenced by the linguistic elements in the contents of some of the selected emails. Of significance in fraud discourse, is the context of religion; because it appeals more to the religious inclination of the intending victims to defraud them. All the emails with religious contents manifest charity services. The email writers deploy in this cognitively based communication religion and its associated activity of charity to seek mutual cognitive environment to perpetrate their dubious act. The following examples further illustrate this.

Example 15

I need a very honest and God-fearing person that can use these funds for God's work (charity) and 15% out of the total funds will be for your compensation for doing this work of God.

Example 16

…I have decided to donate my fund ($2,000,000.00) to you for charitable goals…I want you to use this fund to help the orphanage homes, poor, sick ones in the hospital…Proverbs 19:17: he who gives to the poor lends to the Lord and the Lord will reward such a person for good work.

The co-texts lexicalizing the context of religion and charity services from the foregoing are: "God-fearing", "God's work", "charity", "work of God". "donate", "charitable goals", "fund", "orphanage homes", "Proverbs 19:17", "Lord", "good work". These items are capable of mentally constructing celestial idea within Christian religion from these propositions in the minds of the email recipients, especially those who believe in this system.

Also, two pragmatic strategies related to this context are evocation of religious affinity and evocation of messianic figure. Evocation in this situation implies that the email writers cognitively create these pictures in the emails through their use of language. The evocation of religious affinity relates to the assumption of shared religious belief through which cyber-fraudsters disguise with the intention to defraud the email recipients. This strategy is deployed by using Christianity terms like "God's work", 'charity work", "Your sister in the Lord" and so on. They do this to orient to the religious beliefs of the recipients.

The evocation of messianic figure describes how the email writers project themselves as messiahs, philanthropists to their recipients. They deploy this strategy by mentally creating the figure of generosity to the people they want to defraud. Lexical items depicting this are: "compensation", "reward", "beneficiary", "you will receive 30%" just to mention a few.

6. Conclusion

It has been shown in the analysis and findings that six stages characterise the discourse pattern of email fraud, namely, salutation, discourse initiation, enticing information, mild conscription into business, request and subscription. Through these the email writers (cyber-fraudsters) orient to the context of business and context of religion; thereby hiding under these guises to defraud the email recipients. Through mental representations and propositions evidenced in the contents of the emails, recipients are expected to inferentially process the meanings via contextual assumptions. Meanwhile, pragmatic strategies of adversative and evocation of business idea distinguish context of business, while those of evocations of religious affinity and messianic figure manifest in the context of religion. The preponderance configurations of business and charity instincts in the selected emails align with Young's (2006) findings. This study has also shown that there is a heavy use of polite expressions; importantly deployed to appeal to the positive face want of the recipients to yield easily to their tact. The choice of GSP and cognitive context has significantly made the study to unpack the linguistic configuration of email fraud and track the cognitive undertone of such discourse. The study, therefore, concludes that cyber-fraudsters deploy similarly familiar patterns and contexts evincing strategic persuasive language to defraud their prospective victims. The study complements existing literature on fraud discourse in linguistic scholarship.

References

Barron, Anne. 2006. Understanding spam: a macro-textual analysis. *Journal of Pragmatics* 38(6), 880-904.

Behnam, Biook, Azabdaftari, Behrooz, Hosseini, Ali. 2011. A critical analysis of financial fraud spam in English in terms of persuasive strategies: personalization, presupposition, and lexical choices. *Journal of English Studies: Islamic Azad University, Science & Research Branch* 1(4), 15-26.

Blommaert, Janson. 2005. Making millions. English, indexicality and fraud. *Working Papers in Urban Language & Literacies* 29, 1-24.

Chiluwa, Innocent. 2009. The discourse of digital deceptions and "419" emails. *Discourse Studies* 11 (6), 635-660.

Chiluwa, Innocent. 2010. The pragmatics of hoax email business proposals. *Linguistik Online* 43(3), 32-48.

Chiluwa, Innocent. 2013. Email fraud. *International Encyclopedia of Language and Social Interaction*. Wiley Blackwell & International Communication Association (ICA).

Dechow, Patricia, M. 2011. Predicting material accounting misstatements. In: *Contemporary Accounting Research* 28, 17–82.

Fadipe-Joseph, Olubunmi, A., Titiloye, Emmanuel, O. 2012. Application of continued fractions in controlling bank fraud, *International Journal of Business and Social Science* 3(9), 210-213.

Fetzer, Anita. 2007. *Recontextualising Context*. Amsterdam/Philadelphia: John Benjamins.

Halliday, Michael,.A.K., Hasan, Ruqaiya. 1985. *Language, Context and Text: Aspects of Language in Social-semiotic Perspective*. Oxford: OUP.

Heyd, Theresa. 2008. *Email Hoaxes*. Amsterdam: John Benjamins.

Hua, Tan, K., Abdollahi-Guilani, Mohammad, Zi, Chen, C. 2017. Linguistic deception of Chinese cyber fraudsters. *The Southeast Asian Journal of English Language Studies* 23 (3), 108-122.

Idowu, Abiola. 2009. An assessment of fraud and its management in Nigeria commercial banks. *European Journal of Social Sciences* 10(4), 628-640.

Kerremans, Koen, K., Tang, Yan, Temmenman, Rita, Zhao, Gang. 2005. Towards ontology-based e-mail fraud detection (August, 2005). http://antiplushing.org/APWGplushingactivityReport August2005pdf.

Lan, Li. 2002. Email – a challenge to standard English? *English Today* 16 (4), 23-29.

Leap, Terry. 2007. *The Dynamics of White-collar Crime Ithaca*. NY: Cornell University press.

Odebunmi, Akinola. 2007. Explicatures and implicatures in News magazine editorials: the case of the Nigerian Tell. *Perspective on media discourse.* Rotimi. Taiwo, Akinola. Odebunmi and Akin. Adetunji. (eds.), 84-99.

Orasan, Constantin, Krishnamurthy, Ramesh. 2002. A corpus-based investigation of junk mails. *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, 29–31 May, Las Palmas, Spain, Retrieved from: http://clg.wlv.ac.uk/papers/orasan-02b.

Ong, Walter, J. 1981. *Fighting for life: contest, sexuality, and consciousness.* Ithaca: Cornell University Press; Amherst: University of Massachusetts Press.

Polick M.Y. 2006. What is fraud? available at http://www.wisegeek.com.

Runkehl, Jens, Siever, Torsten and Schlobinsk, Peter. 1998. *Sprache+und+KomW munikation+im+ Internet:+Uberblick+und+Analysen*.! Opladen: Westdeutscher! Verlag.

Sperber, Dan, Wilson, Deirdre. 1986. *Relevance Communication and Cognition*. Oxford: Blackwell.

Young, Michael, R. 2006. *Accounting Irregularities and Financial Fraud: A Corporate Governance Guide*. 3rd edition. Chicago: CCH.

NOTE ON CONTRIBUTORS

M.G. Lalith Ananda is a senior lecturer in Linguistics in the Department of English and Linguistics, University of Sri Jayewardenepura, Colombo, Sri Lanka. He holds a PhD in Generative Syntax from Jawaharlal Nehru University, India (Topic: Clausal Complementation in Sinhala). His research interests are in the area of Sinhala Syntax with particular interest in Topic, Focus, Mood, Modality phenomena. Email address: mlalithananda@gmail.com

Prof. Krzysztof Hejwowski is a translation scholar working at the Institute of Applied Linguistics, University of Warsaw. He has published three books (*Translation: a cognitive-communicative approach*, *Kognitywno-komunikacyjna teoria przekładu* and *Iluzja przekładu*) and numerous articles dealing with translation theory and practice. He has edited or co-edited several volumes of articles devoted to translation problems, notably the four volumes of the *Imago mundi* series. He is particularly interested in translation theory and literary translations. Email address: k.hejwowski@uw.edu.pl

Paweł Dziedziul is a PhD candidate in the faculty of philology at the University of Białystok, Poland, as well as a research assistant and lecturer at the English philology Institute. He also works in the IT industry on software testing and quality assurance. He is a member of the editorial board of The Ludwig von Mises – Europe. His research interests are in the areas of theoretical linguistics, generative grammar, philosophy of language, mind and science, AI, Austrian School of economics. Email address: paweldziedziul@gmail.com

Raoul Karimov, born October 16, 1993 in Chelyabinsk, Russia; graduated from Chelyabinsk State University as a Bachelor of Linguistics in 2014 and as a Master of Linguistics in 2016; currently a PhD student at Chelyabinsk State University. He has completed a Summer School of German Language and Cross-Cultural Communication in 2013 at the University of Bremen, Germany; studied for one year (2017–2018) at the University of Bergen, Norway, under the Russian-Norwegian Study Grants Program. His research interests are: corpus linguistics, applied linguistics, old Germanic languages, and machine learning. Email address: raoul.karimov@hotmail.com

Ezekiel Opeyemi Olajimbiti (PhD) teaches in Elizade University, Ilara Mokin, Ondo State. He is a member of English Scholars' Association of Nigeria (ESAN) and Nigerian Pragmatics Association (NPrA). He specialises in the representation of children in the media. His research interests cover semantics, pragmatics, discourse analysis, sociolinguistics and stylistics. Email address: opebukola56@gmail.com