Papers from the 8th Lancaster University Postgraduate Conference in Linguistics & Language Teaching

Volume 8. Papers from LAEL PG 2013

Edited by

Bárbara-Pamela Olmos-López

Jing Huang

João Almeida

2013

Contents

Introduction to the Volume

The aim of the Postgraduate Conference in Linguistics and Language Teaching is the academic exchange of postgraduate students' research. As every year the conference includes diverse topics in linguistics and language teaching. However, only a selected number of papers are included in the conference proceedings. Volume 8 brings a selection of four articles: *Adjective Stacking and Classification in Northern Sotho: A Southern Bantu Language of South Africa* by Paul Flanagan; *Effects of task repetition on written language production in Task Based Language Teaching,* by Bimali Indrarathne; *Investigating the listening construct underlying listening-to-summarize tasks* by Anchana Rukthong, and *Second language writing development from a Dynamic Systems Theory perspective* by Attila M. Wind. The authors provide interesting and innovative perspectives to doing research in linguistics, and in this volume, three are focused on language teaching and assessment.

Paul Flanagan presents an innovative article in the context of South African language in Sotho. The value of this article lies on the methodology used. Paul skilfully combines field work methods with a corpus approach, which, as he suggests, has been controversial in grammatical study. In his study, he then shows how these methods can be combined to produce a more complete analysis of adjectives. The background of South African languages is shown to be a rich context to analyse and explore language features as he does. This article is part of his doctoral research, and those interested in the area will want to read his thesis.

In the field of language teaching, Bimali Indrarathne centres her research on task repetition for writing production. In her article she provides a useful introductory discussion of literature which takes the reader by the hand to understand how task repetition impacts language learning. She particularly devotes attention to whether task repetition increases fluency, accuracy and complexity by doing a case study. The article provides evidence of the

appropriate choice of methodology, research instruments and the analysis carried out. The careful research procedure shows how task repetition can work not only to increase written performance, but also to transfer knowledge of discourse features. Bimali's article points to some pedagogical implications which benefit the language classroom.

Anchana Rukthong's article moves to the area of language assessment. Her study focuses on the listening construct underlying listening-to-summarise academic lectures. In her literature review, Anchana describes the cognitive processes when performing listening tasks in an academic context in line with the factors that contribute to effective listening. Her pilot study reports a successful attempt that challenges the way language has been assessed. The methodology used were four tests (two oral tasks and two written tasks) which were applied to Thai students living in the UK and belonging to an academic community. Her study sheds light on the process of performing and test measuring; the evolution of her PhD research will certainly provide a richer understanding of assessment.

Attila Wind   concludes this volume with his research on writing development. He centres his study on the lexical and syntactic development of English as a Foreign Language (EFL) learners. He explains the connection of L2 writing development and Dynamic System Theory which allows a clear transition to understand his results. The value of his study is the methodology used, a longitudinal case study over a four-month period. Attila uses a mixed methods design including qualitative and quantitative instruments. The two main features he analyses are the development of lexical and syntactic complexity. This article is a pilot study of Attila's doctoral research which aims to shed light in the process of L2 writing development.

In this introduction, in the name of the editorial committee, I would also like to thank to the manuscript reviewers who kindly donated their expertise and time to reading the papers.

<div align="right">
Bárbara-Pamela Olmos-López
Lancaster University
</div>

Adjective Stacking and Classification in Northern Sotho: A Southern Bantu Language of South Africa

Paul Flanagan
Edge Hill University, UK

Abstract

In this paper, I investigate the nature of complex nominal modification in Northern Sotho, a Southern Bantu language and an official language of South Africa. Adjectives in Northern Sotho have traditionally been analysed as a subclass of nouns, based on morphological similarities between nouns and adjectives. Based on recent work on both Bantu lexical classification and the category 'adjective', I refer to the range of Northern Sotho nominal structures commonly termed 'qualificatives' as adjectives. There has been some limited discussion of the theory that there are universal structures in adjective order across different languages, although sequencing in languages with postnominal adjectives remains under-researched. Using a combination of corpus data and original fieldwork, I provide support for the suggestion that there are patterns in the syntax of complex modification strings which operate on a universal level, above that of individual languages.

1. Introduction

This paper documents an investigation into the structure and organisation of poly-adjectival noun phrases (PNPs, Bache 1978) in Northern Sotho, a Southern Bantu language spoken mainly in South Africa. In section 2, I give some background on the language of Northern Sotho and in section 3, I discuss the phenomenon of 'adjective stacking' and outline critical theory which attempts to explain the ordering of attributive adjectives across different languages. In section 4, I provide an account and justification of my choice of a mixture of corpus and field methods for my study, before giving a detailed overview of the morphology, syntax and semantics of the adjective class in Northern Sotho in section 5. In section 6, I present my results and provide some discussion of these data before offering some conclusions on the nature and structure of complex modification in Northern Sotho.

This study relates not only to the phenomenon of adjective stacking, but also to the wider argument concerning the nature of the adjective class as an independent lexical category. Although major theorists have suggested that not all languages have a word class of adjectives (Jespersen 1961; Rijkhoff 2002: 133; McGregor 2009: 84; Dryer 2013), in recent years, there has been considerable work done to provide a counter-argument to this assumption, most notably by Baker (2003) and Dixon and Aikhenvald (2004). In this paper, I refer to refer to the structures labelled by Poulos and Louwrens (1994) as 'qualificatives' as adjectives, rather than their traditional analysis as a subclass of nouns (Van Wyk 1967). My study also provides evidence which supports Sproat and Shih's suggestion (1988) that complex modification chains show similarities in their internal order across different languages.

2. Northern Sotho: Background

Northern Sotho is a language or language cluster spoken predominantly in the North-East of South Africa by around 4 million people in the Limpopo and Mpumalanga provinces (Paul et al 2013). The language is commonly referred to as Sepedi, which is the dialect from which the standard form was developed by German missionaries in the 19th Century. However, this is only one of 27 dialects of the language (Mokgokong 1966), with other varieties such as Lobedu exhibiting considerable variation from Sepedi (Kotze 2001). There is a high degree of mutual intelligibility between Northern Sotho and other Southern Bantu languages, with Southern Sotho and more noticeably Tswana plausibly constituting something of a linguistic continuum with the language. Lombard et al (1985: 5) suggest that 'some speakers of a Northern Sotho dialect will understand speakers of a Tswana dialect better than they would understand speakers of another Northern Sotho dialect'.

During my work with the language, it became apparent that the nomenclature of the Northern Sotho language was a subject of considerable sensitivity to native speakers of the language. In the Constitution of South Africa (1998, available on *gov.za*, the South African Government website), the language is referred to as 'Sepedi' in its capacity as a designated official language. The African Languages Website (2013) suggests that the language is often 'wrongly referred to as "Sepedi", while in actual fact Sepedi is considered but a dialect of the language "Northern Sotho"'.

The term Northern Sotho is a label created by Western missionaries, who divided up the indigenous languages of the region as much by geography as by genealogy (Mojela 2008). While *sotho* is listed in some grammars of the language as a colour term (brown), none of my informants recognised this usage. In fact, *sotho* is a portmanteau word which

blends the adjectival root *–so* (*black*) with the nominal root *–tho* (person), and hence has an interpretation which essential refers to 'black people' (Mokgoatjane 2013 p/c).

For the purpose of this investigation, I refer to the language as Northern Sotho as this is the name most commonly used in linguistic literature, and is the term used by *Ethnologue* (Paul et al 2013) and the *World Atlas of Language Structures* (Dryer et al 2013).

3. The Order of Attributive Adjectives

The term 'Poly-adjectival Nominal Phrase (PNP)' was developed by Bache (1978) to refer to any noun phrase with more than one modifying adjective, and is the term I use for such structures in this paper. Bache separates adjectives in English PNPs into what he calls functional 'zones of modification', with Mod-I adjectives in the first zone, Mod-II in the second and Mod-III adjectives in the third zone, closest to the head. Mod-I adjectives such as *usual* are said to specify or identify a noun; Mod-II adjectives such as *big* describe or characterise a noun; Mod-III adjectives such as *political* classify or categorise a noun. Therefore, the ordering within a PNP such as *the usual big political issues* can be analysed based on Bache's theory into these three functional zones.

The order of adjectives within the second or 'central zone' (Quirk et al 1985: 1338) has been the subject of considerable investigation by linguists, with many theories proposed which predict adjective order based on the identification and sequencing of a variety of semantic subclasses. Scott (2002: 92) suggests that theorists do not always agree in their observations of 'just how many semantic categories of adjective there actually are', and points out that there is no general agreement on the ordering of such semantic categories. Even a brief glance at some of the more commonly cited semantic order typologies of PNPs shows that a number of patterns consistently occur:

Dixon (1982: 17):

VALUE > DIMESION > PHYSICAL PROPERTY > SPEED > HUMAN PROPENSITY > AGE > COLOUR

Goyvaerts (1968: 27):

QUALITY > SIZE > LENGTH > SHAPE > AGE > COLOUR > NATIONALITY > STYLE > GERUND > DENOMINAL

Quirk and Greenbaum (1973: 404)

GENERAL > AGE > COLOUR > PARTICIPLE > PROVENANCE > NOUN > DENOMINAL

Alexander (1992: 86)

OPINION > SIZE > AGE > SHAPE > COLOUR > ORIGIN/NATIONALITY > PARTICIPLE

Scott (2002: 114)

DETERMINER > ORDINAL NUMBER > CARDINAL NUMBER > SUBJECTIVE COMMENT > EVIDENTIAL > SIZE > LENGTH > HEIGHT > SPEED > DEPTH > WIDTH > WEIGHT > TEMPERATURE > WETNESS > AGE > SHAPE > COLOUR > NATIONALITY/ORIGIN > MATERIAL > COMPOUND ELEMENT > HEAD

It is clear to see from these examples the difference of opinion to which Scott refers. However, it is also noticeable that each ordering features 'age' before 'colour' and 'nationality', with 'size' preceding them in some form (dimension, general) along with some manner of evaluative judgement (quality, value, opinion, subjective comment, general). It is a notion commonly agreed upon that English adjectives denoting subjective properties of a head tend to precede those which are more objective and inherent (Whorf 1937; Teyssier 1968; Quirk et al 1985: 1339; Adamson 2000: 55-6; Wulff 2003). Many theorists combine an explanation of adjective order based on functional and semantic factors, with an appreciation that contextual factors also affect such orders. Greenberg and Srinivasan (2003: 1) note that 'there are some cases where multiple orderings are not only legitimate, but carry different semantic content', and Quirk et al (1985: 1341) suggest that speakers will often place adjectives in an order which 'corresponds to the 'natural' order of recursive qualification',

and which sometimes 'reflects the non-linguistic world'. Often an adjective will be placed earlier in the sequence either to foreground and emphasise it, or even to suggest an adverbial relationship with the following modifier, as in the phrase *the beautiful tall trees*.

While adjective order has been studied to a considerable extent in English, the same cannot be said about other languages. As noted, the adjective class is one which divides linguists as to whether all languages possess such a category, but those languages for which a distinctive adjective class is recognised exhibit considerable cross-linguistic variation in their syntax and morphology. While English has adjectives which appear before the noun (with the exception of occasional French calques and poetic usages), an estimated 62% of languages have adjectives which predominantly appear after the noun (Dryer 2013). Northern Sotho has exclusively postnominal adjectives and this study is intended to investigate to what extent the order of adjectives varies between Northern Sotho and English.

Sproat and Shih (1991: 591) suggest that the ordering of adjectives in Chinese and English is largely very similar, and relates this to the fact that both languages have typically prenominal attributive adjectives. The same authors also note, however, that the order of adjectives should be 'stated in terms of closeness to the head, rather than in terms of linear ordering', and suggest that comparing this feature in languages with different orderings of adjective and noun is more complex than in languages with similar constituent orders in nominal phrases.

Cinque (1994: 87; 2010) compares adjective placement in Romance and Germanic languages, and suggests that ordering is to some extent universal. He notes that two possible orderings of adjectives exist: the same order as English, and the reverse, or mirror-image order of English. Willis (2011: 1807) argues against this observation, claiming that such an analysis is 'grounded in the theoretical context of recent work in comparative syntax'. He suggests that N-raising analyses of noun phrases such as that proposed by Cinque are

insufficient in explaining the nature of PNPs in Welsh, which, as a language with predominantly postnominal adjectives, exhibits both patterns Cinque observes, as well as other more distinctive orders. An analysis of the ordering in Northern Sotho offers a fresh perspective on this issue.

4. The Adjective Class in Northern Sotho

4.1 Morphosyntax

Like other Bantu languages, Northern Sotho is highly agglutinative in nature (Nurse and Philippson 2003: 44) and thus the morphological structure and syntax of attributive adjectives is particularly different to that of English. Adjectives are nearly always postnominal, are separated from the head by a particle and (in the case of central adjectives) an inflectional prefix, both of which express concord with the noun class of the head. Many theorists have referred to the adjective as a subclass of noun (Lombard 1985: 58, Ziervogel 1969, Van Wyk 1967), others refer to adjectival constructions (Prinsloo et al 2013) and Poulos and Louwrens (1994) consider 'qualificatives' as a loose equivalent of the adjective class in English. The usage by the first three authors of the term 'adjectival noun' is influenced by Doke's (1954) seminal work on Southern Bantu which does not recognise adjectives as constituting an independent word class, instead labelling them as nouns based on their morphology. For the purpose of this study, the term 'adjective' is used to refer to the range of structures that Poulos and Louwrens (1994) label 'qualificatives' and which Prinsloo et al (2013) label 'adjectival constructions'.

Adjectives in Northern Sotho are made up of a lexical root (-*golo* 'big') and a concordial morpheme (Lombard 1985: 84) which changes to reflect the class prefix of the modified head. The adjective is separated from the head by a small particle of two or three

letters, which also changes to express concord with the noun class of the head. This particle is most commonly called the qualificative particle (Lombard 1985: 171, Poulos and Louwrens 1994: 91), but is also referred to as the demonstrative (Ziervogel 1969: 58) the adjective prefix and the adjective particle (Poulos and Louwrens 1994: 91). Although the particle is very similar in form to the demonstrative, it performs a different function here and works only as a linker between adjectives and nouns.

> *Monna*   *yo*   *mo-golo*
> N1-man   QP1   CM6-ADJbig
>   'big man'

> *Selepe*   *se*   *se-golo*
> N5-axe   QP5   CM5-ADJbig
>   'big axe'

As well as adjectival nouns in Northern Sotho, there exists another group of modifiers which are commonly referred to as 'relative nouns' (Lombard et al 1985: 59-60), but also 'nominal relatives' and 'pseudo-adjectival constructions' (Prinsloo et al 2013: 76). The equivalent category has been analysed in Tswana and labelled as 'new adjectives' (Creissels 2010). The constructions do not share all the morphosyntactic properties of full adjectival nouns, and constitute a class similar to 'peripheral adjectives' in English (see Quirk et al 1985: 1338). While relative nouns follow the qualificative particle to modify a noun, they have an initial bound morpheme instead of inflecting to express concord with the head:

> *Banna*   *ba*   *bohlale*
> N2-men   QP2   RELNclever   (Ziervogel 1969: 55)

Enumerative nouns (Poulos & Louwrens 1994), constitute a very small class of words with only four members, which are similar to adjectival nouns. While they usually follow the subject concord and do not always take a concordial morpheme, they can also function with the qualificative particle either with (like adjectival nouns) or without (like relative nouns) the concordial morpheme:

Metše	*ye*	*me-šele*
N4-villages	QP4	CM4-strange

Metše	*ye*	*šele*
N4-villages	QP4	strange

Metše	*e*	*šele*
N4-villages	SC4	strange

      'foreign villages'   (Poulos & Louwrens 1994: 112)

When more than one adjective is used to modify a noun at the same time, the qualificative particle is repeated for each adjective:

Mahlo	*a*	*ma-botse*	*a*	*ma-tsothwa*
N6-eyes	QP6	CM6-ADJbeautiful	QP6	CM6-ADJbrown

      'beautiful brown eyes'   (De Schryver 2013: 110)

Selo	*se*	*se-ngwe*	*se*	*se-kaone*
N7-thing	QP7	CM7-ADJother	QP6	CM6-ADJbetter

      'another better thing'   (*Pretoria Sepedi Corpus*)

Adjectival nouns can be combined in this way with relative nouns and enumerative nouns, as evidenced by the following examples (from De Schryver 2013: 110-115):

Dintšhi	*tše*	*telele*	*tše*	*boleta*
N8-eyelash	QP8	CMØ-ADJtall	QP8	RELNsoft

      'long soft eyelashes'

Mathebo	*a*	*mabotse*	*a*	*dinkwe*
N6-skins	QP6	CM6-ADJbeautiful	QP6	RELNleopard

'beautiful leopard skins'

4.2 Semantics

The adjectival category constitutes a fairly small and relatively closed class in Northern Sotho. Segerer (2008: 1) notes that the nature of adjective classes in African languages is well documented and suggests that nearly all members Niger-Congo family have a 'small, closed class [in which] the number of items ranges from 2 to more than 100'. This description applies to that of adjectival nouns in Northern Sotho, in which the number of items varies slightly from one source to another. Lombard (1985) lists 30 adjectival bases, while Ziervogel (1969) only names 26. Mphasha (2010) names 33, while Poulos and Louwrens (1994) name 35.

In addition to the roots noted in these sources, De Schryver (2013: 446) suggests that –*pinki* 'pink', also functions as an adjectival noun, a borrowing which opposes the notion that Bantu adjective classes tend to be closed. Similarly, my own study of the *Pretoria Sepedi Corpus* reveals that the adverb *kaone* 'better' (see example on previous page) can also be used as an adjectival noun, a fact confirmed by my native speakers. Overall, it is possible to identify as many as 44 adjectival roots in Northern Sotho, which suggests that it has a fairly large class for a Niger-Congo language, and one which is potentially more open than others.

The adjectives of Northern Sotho can be divided into semantic subclasses as follows:

NUMERAL	SPECIFIER	COLOUR	SIZE	AGE/GENDER	EVALUATIVE
-tee (one)	*-ngwe* (other)	*-hubedu* (red)	*-golo* (big)	*-fsa* (new, young)	*-botse* (beautiful)
-pedi (two)	*-bjang* (what	*-sehla* (grey)	*-nyane* (small)	*-tala* (old)	*-šoro* (cruel, vicious)
-raro (three)	*-bjalo* (such, so)	*-tala* (blue)	*-telele* (tall)	*-tona* (male)	*-šele* (strange)
-ne (four)	*-kaaka* (this/so big)	*-so* (black)	*-koto* (thick)	*-tshadi* (female)	*-thata* (hard)
-hlano (five)	*-kaakang* (how big?)	*-šweu* (white)	*-kopana* (short)		*-be* (bad, evil)
-ntši (many)	*-fe* (which?)	*-tsothwa* (brown)	*-sese* (thin)		*-bose* (nice, tasty)
-kae? (how many?)		*-pinki* (pink)	*-tona* (big)		*-kaone* (better, best)
-nyane (few)		*-tilo* (black &	*-kgopo* (crooked)		*-borethe* (smooth)
		-khunou (reddish-			*-boleta* (nice, tasty)

Figure 1: Semantic Subclasses of Adjectives in Northern Sotho

As shown above, adjectives in Northern Sotho do not occupy the same semantic space as their English counterparts. Numerals, as in other Bantu languages, are adjectives in Northern Sotho, and there are also question words which work in the same way, such as *banna ba bakae?* 'how many men?'. My fieldwork suggests that even those which are also adjectives in English have particular distinctions in their usage and interpretation. Adjectives which denote size and age are very closely linked in Northern Sotho. This means that many phrases

which make reference to both size and age might seem perfectly natural in English but cannot be collocated without a suggestion of contradiction or tautology in Northern Sotho:

*monna	yo	monyane	yo	motala
N1-man	QP1	CM1-ADJsmall	QP1	CM1-ADJold
		'little old man'		

*basetsana	ba	batelele
N2-girls	QP2	CM2-ADJtall
		'tall girls'

*kgomo	ye	nnyane
N9-cow	QP7	CM7-ADJsmall
		'small cow'

*mošemane	yo	mofsa
N1boy	QP1	CM1-ADJyoung
		'young boy'

For the first two noun phrases, speakers of Northern Sotho would consider a strong sense of contradiction in the expression of qualities of size and age. In the first example, the adjective *nyane* 'small' is seen as contradictory to the adjective *tala* 'old'. While the phrase 'a little old man' is a common collocation in English, the Northern Sotho term *nyane* refers not just to something which is small in size, but also has a strong indication of being young in age. Similarly, *kgomo ye nnyane* has a literal translation of 'a small cow', but this term in Northern Sotho refers more to a young cow, more commonly denoted by the noun *namane* 'calf'. Similar problems arise from the terms *basetsana ba batelele* and *mošemane yo mofsa*. *Basetsana* 'girls' are considered to be small and to describe them as 'tall' sounds strange to a native speaker, while to describe a *mošemane* 'boy' as 'young' sounds like a tautology.

While English adjectives can describe the property of an object in relation to other objects of its kind, my informants' comments suggest that Northern Sotho adjectives are more absolutive and there are more complex collocational restrictions, both with heads and with other modifiers. Cinque (2010: 10) refers to these kinds of modification as 'relative' and 'absolute' readings, citing *a big tank* as an example of the former. My research strongly suggests that Northern Sotho adjectives are far more arbitrarily absolute than those in English, which are more flexible. The reflections elicited in my fieldwork are supported by the fact that no examples such as those discussed in the preceding paragraph appear in my corpus data.

Another interesting feature to note here is the variability in the interpretation of certain adjectives such as *–fsa, -telele* and *–tona*. Mphasha (2010: 23) suggests that for some adjectives, the meaning is selected by the head. The adjective *–fsa* can mean 'young' or 'new' depending upon the animacy of the modified head. If a snake or a road were modified by *–telele*, the interpretation would be 'long', while for a person it would be 'tall'. The adjective *–tona* is particularly interesting, as it can mean 'right', 'male' or 'big', depending upon the nature of the modified head. Consider the following examples:

Kgomo   *ye*   *tona*
N9-cow   QP9   CMØ-ADJ-male
     'a male cow'

Seatle-ng   *se*   *setona*
N7-Hand-LOC QP7 CM7-ADJ-big
     'in the right hand'

Mello   *ye*   *mentši*   *ye*   *metona*
N4-fires   QP4 CM4-ADJmany   QP4   CM4-ADJ-big
     'many big fires'

The adjective *–tona* is quite rare in the *PSC* but based on my fieldwork, it seems to be most commonly used to refer to animals rather than people, with such a usage suggesting disrespect in a similar way to the adjective *bullish* might in English. It tends to refer to the alpha male in a group of animals and hence also has the interpretation of being of a large size. De Schryver (2013: 236) suggests that this term can also mean 'important'. This interpretation, along with the usage to refer to the right hand or the right arm, seems to be the result of a semantic drift based on an association with power and strength. Interestingly, the adjective meaning 'left', *–tshadi* also denotes femininity (Prinsloo 2009: 169), suggesting a link between gender and the terms for right and left. This etymology is not quite clear, but it is reasonable to say that Mphasha's (2010: 23) suggestion that meaning is distinguished by head is particularly salient here.

5. Methodology - mixed method approach

For this investigation, I have adopted a mixed-method approach which combines the use of data from the *Pretoria Sepedi Corpus* with original fieldwork in the form of questionnaires with 30 native speakers, all of whom were teachers in the Sekhukhuneland area. The questionnaires are primarily quantitative in nature, but also allow participants to speculate as to what factors affect the order in which they choose to place adjectives in PNPs. The validity of an approach which combines corpus and field methods is widely attested, with Biber et al (1998: 9) suggesting that 'corpus-based analysis should be seen as a complementary approach to more traditional approaches' and that corpus studies alone are not always sufficient to investigate or explain fully, a linguistic phenomenon. McEnery and Wilson (2007: 30) suggest that corpus data allow us to investigate a language through a sample which is 'maximally representative' and which 'provides us with as accurate a picture as possible of the tendencies' of the grammar of a language.

Newman and Ratliff (2001: 49) suggest that it is important to work with native speakers of a language if a researcher wishes to gain valuable, realistic data. They suggest that native speakers 'have background knowledge of context that an outsider may lack' and that 'good speakers have a sense of the difference between not only what is grammatical and what is ungrammatical but also of what is said and what could be said but is not'. While a corpus study provides us with information on commonly-appearing structures, it is not always possible to speculate on the respective grammaticality of structures which do not appear in the data set. Newman and Ratliff (2001: 51) stress this suggestion and add that native speaker informants can offer judgements beyond the realm of a pure corpus-based study. They note that 'some speakers are especially sensitive to nuances of style and register, able to point out the effects of lexical and grammatical choices'.

My corpus data include over 300 Northern Sotho PNPs which feature a range of central and peripheral adjectives. A corpus search was performed of all structures in which a noun was followed by two or more 'adjectives', which included both 'adjectival nouns' and 'relative nouns'. Therefore, all noun phrases studied had one of the following structures:

NOUN-ADJN-ADJN (-ADJN/RELN … and so on)
NOUN-ADJN-RELN (-ADJN/RELN … and so on)
NOUN-RELN-ADJN (-ADJN/RELN … and so on)
NOUN-RELN-RELN (-ADJN/RELN … and so on)

The questionnaire featured 40 questions in which participants were asked to select a preferred ordering for a variety of PNPs in Northern Sotho. A large range of combinations of different semantic types was selected, and the patterns were compared with the corpus data to identify whether the trends were comparable, and to provide a broader, varied and more reliable data set on which to base conclusions. For each question, a head noun was supplied,

along with a minimum of two adjectives which were presented in brackets in alphabetical order:

Dikgabo ……………………………………………………………… *(be/ nyane/ tala/ tsothwa)*
(Monkeys)   (evil/ brown/ small/ old)

Participants were asked to construct the sentence as they felt they would say it in natural language usage. The omission of concordial morphemes and qualificative particles in the question format was also instructive in this measure. In most cases, more than one participant completed the questionnaire simultaneously, and I encouraged them to discuss which order they felt most natural with a view to getting reliable, vernacular data. The data were then analysed, with the number of each possible sequence of adjectives represented as a percentage. In some cases, there is a clearly preferred order, while in others a range of possible orders exist.

6. Results and Discussion

6.1 Corpus Data

As noted above, the corpus data contains over 300 tokens featuring all combinations of two or more 'adjectives', under which umbrella term I refer to both adjectival nouns and relative nouns (Lombard 1985). The results display a number of trends. The most commonly occurring adjective is *–ngwe* 'other' which appears in nearly half of all PNPs in my data set. It appears closest to the head in 91% of tokens, with the remaining 9% most likely marked orders employed to foreground the other adjective. Even numerals appear further from the head than *–ngwe*. Consider the following table:

Collocate	Sem-type	ngwe-X	X-ngwe	Total	% initial
-golo	size	17	3	20	85
-bedi	num	15	2	17	88
-ntši	spec/num	14	2	16	88
-botse	eval	9	0	9	100
-fsa/swa	age	6	0	6	100
-be	eval	5	1	6	83
-nyane	size	6	0	6	100
-raro	num	3	0	3	100
-hlano	num	3	0	3	100
-šweu	col	3	0	3	100
-kaone	eval	3	0	3	100
-so	col	3	0	3	100
-telele	size	1	1	2	50
-ne	num	1	0	1	100
TOTAL		89	9	98	91%

Figure 2: Collocations with –ngwe 'other'

After *–ngwe*, the adjectives most commonly placed closest to the noun are numerals. The most commonly-occurring numeral is *–pedi* 'two', which appears closer to the head than other collocates in 94% of modification strings which do not include *–ngwe*. The other numerals, including *–ntshi* 'many', appear closest to the head in the majority of tokens. After numerals, adjectives of size tend to appear closer to the noun than adjectives from other semantic subclasses. Of these adjectives, *-golo* 'big' is by far the most common, and tends to precede other adjectives of size when they are combined. This is most likely because it is a more general and subjective descriptor, less specific than *–telele* 'tall' or *–koto* 'thick'. Whorf (1937: 93, see also Adamson 2000: 55-6 and Wulff 2003 on this) suggests that subjective judgements tend to precede those which are more objective, and this tendency seems to reflect Whorf's comments. The following table shows the percentage of tokens in which adjectives denoting size appear closest to the head:

	with –ngwe	with numeral	with other size	with colour	with other
-golo	15%	21%	83%	56%	100%
-telele	50%	20%	38%	83%	50%
-nyane	0%	0%	n/a	100%	n/a
-koto	n/a	0%	25%	100%	100%
-šese/-kopana	n/a	0%	33%	100%	n/a
TOTAL	15%	19%	50%	75%	88%

Figure 3: Size Adjectives in String-initial Position

It easy to deduce from these figures that there is a clear pattern in the order of adjectives, which is associated with semantic subclasses. Size adjectives precede colour adjectives and other more generally evaluative adjectives in a high proportion of cases, and follow numerals and *–ngwe* in similar proportion. Due to the fact that colour terms and evaluative adjectives are collocated in very few tokens in the data, it is difficult to provide an extensive theory as to their respective positions. However, in the eight tokens of such structures which exist, colour terms follow evaluative adjectives in each and every example, allowing a moderate deduction that this is the normative ordering. Additionally, adjectival nouns appear closer to the head than relative nouns in 92% of combinations. In light of this, the most common order of adjectives appearing in the *Pretoria Sepedi Corpus* is as follows:

-NGWE > NUMERAL > DIMENSION > EVALUATIVE > COLOUR > RELATIVE

This ordering to a large extent reflects the order in which adjectives are placed in English PNPs and is further explored and tested in my fieldwork.

6.2 Fieldwork

All field-based data were collected during a field trip to the Sekhukhune region of Limpopo, South Africa, which is considered to be the 'birthplace' of Northern Sotho, as it is the dialect area on which the first missionaries based the standardised form of the language (Mojela 2008). My questionnaire requires speakers to create noun phrases from a nominal head and a selection of adjectival roots. As well as exploring the conventions which seem to exist based on my corpus data, I was able to gain valuable, more qualitiative insights into the usage and interpretation of adjective strings, such as those discussed in 4.2. The quantitative data elicited in my fieldwork largely reflect the conclusions drawn from my corpus data in 6.1, with numerals and size adjectives being selected closest to the head noun in a high proportion of questions. The questionnaire data was also used to determine the order of colour and evaluative adjectives, as well as the extent to which nominal and verbal relatives are restricted to the slot further from the head than adjectival nouns. The table in *Figure 4* illustrates the extent to which numerals appear in string-initial position:
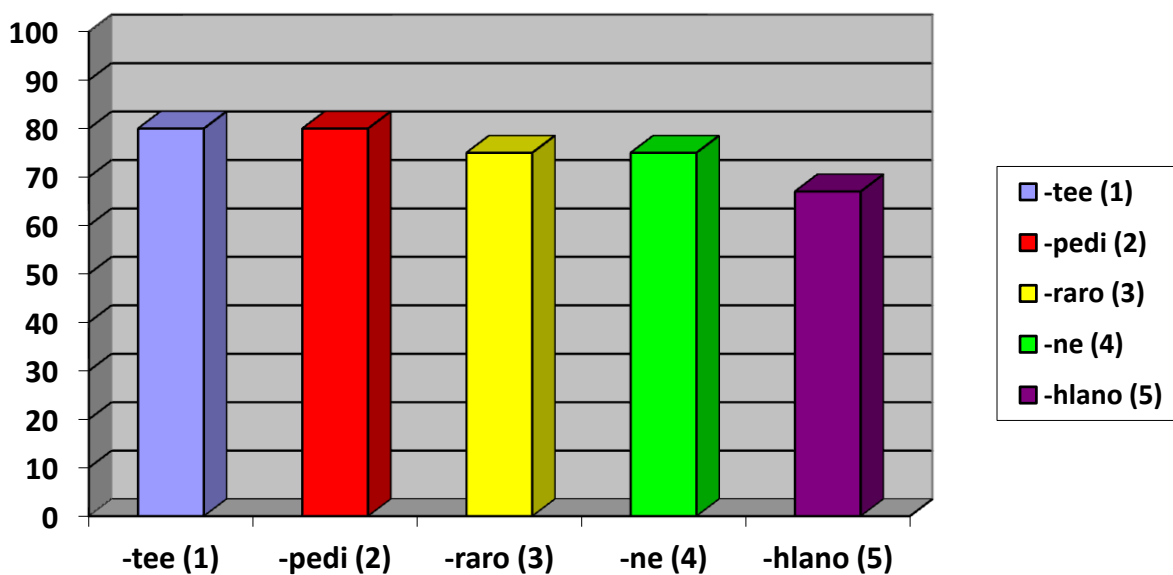

Figure 4: Graph illustrating percentage of tokens with numeral in string-initial position

A major element I wished to investigate was whether orders were fixed or could be changed to emphasise one or another quality. In particular, I wanted to know if numerals and *–ngwe* could be preceded by adjectives typically appearing later in strings from my corpus data. These are properties of PNPs which are very difficult to analyse through corpus-methods alone and strongly justify a mixed-method approach. While the normative ordering reflects that in the corpus examples, participants were generally in agreement that this order could be reversed to indicate emphasis or contrast.

Hence, the following two phrases are both possible:

Dieta	*tše*	*pedi*	*tše*	*ntsho*
N6-shoes	QP6	CM6-ADJtwo	QP6	CM6-ADJblack

'two black shoes'

Dieta	*tše*	*ntsho*	*tše*	*pedi*
N6-shoes	QP6	CM6-ADJblack	QP6	CM6-ADJtwo

'the black two shoes'

While the first example is the more normative ordering for *two black shoes*, the second example is also possible, but with a marked emphasis on the colour. In this context, this is most likely to suggest contrast with shoes of other colours, such as when suggesting one prefers *the black two shoes* to *the brown two shoes*. This sounds a little awkward in English and not all participants agreed that such a change is possible in Northern Sotho. However, 20% of participants felt this was the more natural-sounding order, which strongly suggests that there is a degree of reversibility in such structures.

Another interesting finding was that certain adjectives do not naturally appear in modification strings. The interrogative-type adjectives *–kae* 'how many' and *–bjang* 'what kind of?' are the most notably resistant to co-ordination. The former does not seem to

23

collocate effectively with *–ngwe* while the using the latter in conjunction with other adjectives is particularly problematic. Participants asked to describe a vase using *–bjang*, *-so* and *–botse* found this very difficult. While it is possible to ask the question 'what kind of beautiful black vase would you like?' in English, native speakers generally agreed that such a compound sounded unnatural in Northern Sotho, with the other adjectives sounding like a response to the question posed by the use of *–bjang*.

The data obtained through my fieldwork largely reflect the order suggested by the corpus data, though I also investigated the combination of evaluative adjectives with colour terms in six different phrases. Overall, the evaluative adjectives precede colour terms in 75% of responses, lending support to my suggestion based on corpus data that this is the normative order. Where colour terms precede evaluative adjectives, this usually reflects contexts in which the colour is intrinsic to the head, while the evaluative adjective is more notional and transitory. Consider the following examples:

mae   a   manyane   a   mašweu   a   mabose
N4-eggs QP4 CM4-ADJsmall QP4 CM4-ADJwhite QP4 CM4-ADJsweet
                    'delicious small white eggs'

mahlo   a   mašweu   a   magolo
N4-eyes QP4 CM4-ADJwhite QP4 CM4-ADJbig
                    'big white eyes'

In both of these phrases, the colour term *–šweu* 'white' is an intrinsic and important quality of the respective head. In the first example the colour term has a strong classificatory function (Warren 1984), in which it sub-classifies the head as being a 'white egg' rather than a 'brown egg'; in the second, it has an emphatic function. As all human eyes are by definition white, the usage of this colour term suggests that the whiteness is particularly striking and

noticeable, and is hence foregrounded by being moved closer to the noun, raised above the dimension adjective –*golo*.

One notable trend in my elicited data is that phrase structure becomes less predictable as more adjectives are added. In questions with three of more adjectives, it was often the case that no ordering was considered the preferred order by a considerable majority of participants. In the table in *Figure 5*, the range of responses are illustrated for question 26, for translations of the English phrase *small soft brown loaves*. While *mannyane* 'small' appears closest to the head in the responses given by two thirds of participants, a more even spread was recorded for the ordering of the colour term *matsotho* 'brown' and the nominal relative *boleta* 'soft'. This illustrates not only that the order of adjectives becomes more complex as the number of adjectives in a string increases, but also that nominal relatives are not constrained to appearing further from the head than strings of more central adjectives.

26	Marotho a mannyane a matsotho a boleta	size-col-rel	10	33
	Marotho a mannyane a boleta a matsotho	size-rel-col	7	23
	Marotho a boleta a mannyane a matsotho	rel-size-col	4	13
	Marothwana a matsotho a boleta	Hsize-col-rel	3	10
	Marotho a matsotho a mannyane a boleta	col-size-rel	2	7
	Marotho a boleta a matsotho a mannyane	rel-col-size	1	3
	Marotho a matsotho a boleta a mannyane	col-rel-size	2	7
	(no answer)	(n/a)	1	3

Figure 5: Responses to Question 26

This article provides an overview of the findings of both my corpus data and fieldwork, which are discussed in more detail in my PhD thesis (Flanagan 2014, Ch. 7). The quantitative results relating to all 42 questions are included in *Appendix 1*.

7. Summary Remarks and Conclusions

The data from the *Pretoria Sepedi Corpus* and those obtained through questionnaires with native speakers both suggest that the ordering of adjectives in Northern Sotho PNPs follows a number of trends. Adjectival nouns typically precede relative nouns, and adjectives of dimension follow quantifiers (*–ngwe* and numerals) but are placed closer to the head than adjectives of evaluation and colour terms. This ordering is very similar to the order in which English adjectives are placed (Scott 2002, Dixon 1982). There is evidence to suggest that certain collocations of adjectives in Northern Sotho are problematic or even ungrammatical due largely to semantic restrictions.

There are, however, factors governing the order of attributive adjectives which go beyond the sentence-level, with such pragmatic considerations as emphasis, focus and contrast all being relevant. This suggests that, as well as proposing an arbitrary, normative order in which Northern Sotho adjectives are placed, we can study the respective force of individual adjectives within a string based on the extent to which these conventions are observed. When some property of a noun is to be emphasised (often its colour), the adjective denoting this quality is often raised upwards in the phrase and closer to the head.

Adjective order in Northern Sotho is similar to English in a number of ways. It is governed by collocational and syntactic restrictions, with some orders being fixed and arbitrary, while others are variable. Changing the order of adjectives in a string can have a variety of effects: It can change the meaning of the phrase; it can shift the focus of the phrase to emphasise one adjective over another; and it can even make the phrase ungrammatical. In unmarked orders, the sequence of adjectives is to a large extent similar to English and follows the order:

QUANTIFIER > DIMENSION > EVALUATIVE > COLOUR > RELATIVE

This order reflects that proposed by the theories in section 3 of this article, in so far as numerals (determiners) precede adjectives in English, and size adjectives and derived adjectives appear in string-initial and string-final positions respectively. Relatives occupy a similar position in Northern Sotho as derived adjectives do in English; their close correspondence with nominal and verbal forms is similar in many ways semantically and morphologically to both denominal and deverbal adjectives in English, and this correspondence of placement within a string suggests a level of equivalence between the two subclasses. The potential for evaluative/ subjective adjectives to precede size adjectives seems to be less considerable in Northern Sotho than in English, but the position of colour adjectives toward the end of a string reflects the position of colour terms in English-based order theories.

The fact that the structure of complex modification strings in Northern Sotho is similar to English is most interesting. As the syntactic and morphological structures of the two languages are considerably different, the factors which govern the ordering of adjectives on a cross-linguistic level cannot legitimately be related to grammatical form. The correlation between the positions assigned to the respective semantic subclasses of adjectives in English and Northern Sotho strongly suggests that there are universal syntactic structures which operate in accordance with these semantic subclasses. The possible variability in many orderings supports the notion that the ordering of adjectives is dependent upon discourse features such as focus and emphasis, but the overwhelming evidence here suggests that semantics is the dominant factor which governs adjective ordering at a universal level.

References

Adamson, S. (2000) 'A Lovely Little Example: Word Order Options and Category Shift in the Premodifying String' In Fischer, O., Rosenbach, A. & Stein, D. [Eds.] *Pathways of Change. Grammaticalisation in English* Amsterdam: John Benjamins 39-66

Alexander, L.G (1992) *Longman English Grammar* London: Longman

Bache, C (1978) *The Order of Pre-modifying Adjectives in Present-Day English* Odense University Press

Baker, M.C (2003) *Lexical Categories: Verbs, Nouns and Adjectives* Cambridge University Press

Biber, D, Conrad, S & Reppen, R (1998) *Corpus Linguistics: Investigating Language Structure and Use* Cambridge: CUP

Cinque, G, (1994) 'On the Evidence for Partial N-Movement in the Romance DP'in Cinque, G, Koster, J, Pollock, J-Y, Rizzi, L, Zanuttini, R [Eds.] *Paths Towards Universal Grammar: Studies in Honor of Richard S. Kayne* Washington: Georgetown University Press pp. 85-110

Cinque, G (2010) *The Syntax of Adjectives: A Comparative Study* Massachusetts Institute of Technology

De Schryver, G-M (Ed) (2013) *Oxford Bilingual School Dictionary: Northern Sotho and English / Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.

Dixon, R.M.W (1982) *Where Have All the Adjectives Gone?, and Other Essays in Semantics and Syntax* Berlin: Mouton de Gruyter

Dixon, R.M.W & Aikhenvald, A.Y (2004) *Adjective Classes: A Cross-linguistic Typology* Oxford University Press

Doke, C.M (1954) *Southern Bantu Languages* London: Oxford University Press

Dryer, M.S (2013) 'Order of Adjective and Noun' in Dryer, M.S. & Haspelmath, M [Eds.] *The World Atlas of Language Structures Online* Leipzig: Max Planck Institute for Evolutionary Anthropology (Available online at http://wals.info/chapter/87) [24/1/14]

Flanagan, P.J. (2014) 'A Cross-linguistic Investigation of the Order of Attributive Adjectives' [PhD Thesis] Edge Hill Univerity, UK

Goyvaerts, D.L. (1968) 'An Introductory Study on the Ordering of a String of Adjectives in Present Day English' *Philologia Pragensia 11*, 12–28

Greenberg, A & Srinivasan, P (2003) 'A Survey of Algorithms to Properly Order Adjective Trigrams' Stanford University- available online from nlp.stanford.edu/courses/cs224n/2003/fp/praveens/finalproj.pdf [18/1/14]

Jespersen, O (1961) *A Modern English Grammar on Historical Principles: Part II Volume I* [4th Ed] London: George Unwin & Allen

Kotze, A.E (2001) 'Prominent Features of Lobedu Nouns and Pronouns' *South African Journal of African Languages* 21: 1 (available in html form only)

Lombard, D.P, Van Wyk, E.B & Mokgokong, P.C (1985) *Introduction to the Grammar of Northern Sotho* Pretoria: Van Schaik

McGregor, W (2009) *Linguistics: An Introduction* London: Continuum

Mojela, V.M (2008) 'Standardization or Stigmatization? Challenges of Confronting Lexicography and Terminology in Sesotho sa Leboa' *Lexikos* 18 119-130

Mokgokong, P.C (1966) *A Dialect-geographical Survey of the Phonology of the*

Mphasha, L.E (2010) *Adjectives in Northern Sotho: A Critical Overview of the Aspects of Theoretical Syntax, Morphology and Lexical Semantics of Words Which Modify Substantives* Berlin: Verlag Dr Müller

Newman, P & Ratliff, N [Eds.] (2001) *Linguistic Fieldwork* Cambridge: CUP

Nurse, D & Philippson, G (2003) *The Bantu Languages* London: Routledge

Paul, L.M, Simons, G.F & Fennig, C.D [Eds] (2013) *Ethnologue: Languages of the World* [17th Ed] Dallas: SIL Internations (Available Online http://www.ethnologue.com) [24/1/14]

Poulos, G & Louwrens, L.J (1994) *A Linguistic Analysis of Northern Sotho* Pretoria: Via Afrika

Prinsloo, D.J, Kriel, T.J and Sathekge, B.P (2009) *Popular Northern Sotho Dictionary* [6th Ed] Cape Town: Pharos

Prinsloo, D.J, Van Wyk, E.B, Groenewald, P.S, Kock, J.H.M & Taljard, E (2013) *Northern Sotho for First Years* [8th Ed] Pretoria: Van Shaik

Quirk, R & Greenbaum, S (1973) *A Student's Grammar of the English Language* Essex: Longman

Quirk, R, Greenbaum, S, Leech, G & Svartvik, J (1985) *A Comprehensive Grammar of the English Language* London: Longman

Rijkhoff, J (2002) *The Noun Phrase* New York: Oxford University Press

Scott, G.J 'Stacked Adjectival Modification and the Structure of Nominal Phrases' in Cinque, G [Ed] (2002) *Functional Structure in DP and IP* New York: Oxford University Press

Segerer (2008) 'Closed Adjective Classes and Primary Adjectives in African Languages' *Langage, Langues et Cultures d'Afrique Noire (LLACAN)* (Available online at http://halshs.archives-ouvertes.fr/halshs-00255943) [24/1/14]

Sproat, R & Shih, C (1988) 'Prenominal Adjective Ordering in English and Mandarin' in *Proceedings of NELS* 465-489

Sproat, R, Shih, C, (1991) 'The Crosslinguistic Distribution of Adjective Ordering Restrictions' in Georgopoulos, C, Ishihara, R [Eds.] *Interdisciplinary Approaches to Language: Essays in Honor of S. Y. Kuroda*. Kluwer, Dordrecht, pp. 565-594

Teyssier, J. (1968) 'Notes on the Syntax of the Adjective in Modern English'. *Lingua* 20 225-249 Amsterdam: North Holland Publishing Co.

Warren, B (1984) *Classifying Adjectives* Gothenburg: Humanities Press

Willis, D (2011) 'Against N-Raising and NP-Raising Analyses of Welsh Noun Phrases' *Lingua* 116(11): 1807 - 1839

Van Wyk, E.B (1967) 'Northern Sotho' in *Lingua* 17 230-261

Whorf, B.L. (1937) 'Grammatical Categories'. In Carroll, J.B [ed] (1956) *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf* Massachusetts: MIT Press 87-101

Wulff, S. (2003) 'A Multifactorial Corpus Analysis of Adjective Order in English'. *International Journal of Corpus Linguistics* 8 (2) 245-282

Ziervogel, D, Lombard, D.P & Mokgokong, P.C (1969) *A Handbook of the Northern Sotho Language* Pretoria: J.L Van Schaik Ltd

The African Languages Website www.africanlanguages.com [13/4/2013]

'The South African Constitution' The South Africa Government Website www.info.gov.za/documents/constitution/1996/a108-96.pdf [13/4/2103]

Appendix 1: Quantitative Data from Sekhukhune Questionnaires.

Q	Phrase	Order	#	%
1	**Pere e tee ye ka(aka)**	**num-spec**	**24**	**80**
	Pere ye ka(aka) e tee	spec-num	5	17
	Ye ka(aka) pere e tee	spec-H-num	1	3

'One horse this big'

2	**Dieta tše pedi tše ntsho**	**num-col**	**24**	**80**
	Dieta tše ntsho tše pedi	col-num	6	20

'Two black shoes'

3	**Maboto a mararo a makoto**	**num-size**	**22**	**73**
	Maboto a makoto a mararo	size-num	8	27

'Three thick walls'

4	**Dipotšišo tše nne tše thata**	**num-eval**	**22**	**73**
	Dipotšišo tše thata tše nne	eval-num	8	27

'Four difficult questions'

5	**Dintlo tše tlhano tše mpsha**	**num-age**	**20**	**67**
	Dintlo tše mpsha tše tlhano	age-num	10	33

'Five new houses'

6	**Mpša ye kgolo ye tshothwa**	**size-col**	**23**	**77**
	Mpša ys tshothwa ye kgolo	col-size	7	23

'Big brown dog'

7	**Hempe ye botse ye tala**	**eval-col**	**23**	**77**
	Hempe ye tala ye botse	col-eval	7	23

'Beautiful green shirt'

8	**Polo ye mpsha ye khubedu**	**age-col**	**22**	**73**
	Polo ye khubedu ye mpsha	col-age	8	27

''New red ball'

9	**Dikuane tše tshothwa tše ntši**	**col-spec**	**16**	**53**
	Dikuane tše ntši tše tshothwa	spec-col	14	47

Many brown hats

10	Phiri ye kgolo ye mpe	size-eval	23	77
	Phiri ye mpe ye kgolo	eval-size	7	23

'Big bad wolf'

11	Kgomo ye nnyane ye tona	size-gender	16	53
	Kgomo ye tona ye nnyane	gender-size	7	23
	Namane ye tona	Hage-gender	4	14
	Namane ya pholwana	Hage-gendersize	2	7
	(no answer)	*(no answer)*	1	3

'Small male cow'

12	Basadi ba bangwe ba batelele	spec-size	28	93
	Basadi ba batelele ba bangwe	size-spec	2	7

'Other tall women'

13	Beine ye bose ye mpsha	eval-age	16	53
	Beine ye mpsha ye bose	age-eval	14	47

'Delicious new wine'

14	Modiro wo mothata wo mobjalo	eval-spec	11	37
	Modiro wo mobjalo wo mothata	spec-eval	8	27
	Modiro wo mothata bjalo	eval-spec/adv	5	17
	Modiro wo mothata wo bjalo	eval-rel	3	10
	Modiro wo bjalo wo mothata	rel-eval	2	7
	Modiro woo o mothata	dem-eval	1	3

'Such difficult work'

15	(ke) Dimpša tše tala tše kae	age-spec	18	60
	(ke) Dimpša tše kae tše tala	spec-age	12	40

'How many old dogs?'

16	Motho yo mongwe o tee yo mobe	spec-eval-num	10	33
	Motho yo mongwe yo mobe	spec-eval	10	33
	Motho yo mongwe yo mobe o tee	spec-eval-enum	5	17
	Motho o tee yo mobe yo mongwe	num-eval-spec	3	10
	Motho yo mobe yo mongwe	eval-spec	1	3

	Motho o tee yo mongwe yo mobe	num-spec-eval	1	3

'One other bad person'

17	**Mae a mannyane a mašweu a mabose**	**size-col-eval**	**14**	**47**
	Mae a mabose a mannyane a mašweu	eval-size-col	5	17
	Mae a mašweu a mannyane a mabose	col-size-eval	4	13
	Mae a mannyane a mabose a mašweu	size-eval-col	3	10
	Mae a mašweu a mabose a mannyane	col-eval-size	2	7
	Mae a mabose a mašweu a mannyane	eval-col-size	1	3
	Mayana a mabose a mašweu	Hsize-eval-col	1	3

'Delicious small white eggs'

18	**Sebjanatsopa se sebjang se sebotse se seso**	**spec-eval-col**	**12**	**40**
	(various)	(no answer)	6	20
	Sebjanatsopa se sefe se sebotse se seso	spec-eval-col	3	10
	Sebjanatsopa se seso se sebotse se sebjang	col-eval-spec	3	10
	Sebjanatsopa se sebjang se seso se sebotse	spec-col-eval	2	7
	Sebjanatsopa se seso se sebjang se sebotse	col-spec-eval	2	7
	Sebjanatsopa se sebotse se seso se sebjang	eval-col-spec	1	3
	Se sebjang sebjanatsopa se sebotse se seso	spec-H-eval-col	1	3

'What kind of beautiful black vase?'

19	**Makako a mabedi a makoto a mašweu**	**num-size-col**	**12**	**40**
	Makako a mabedi a mašweu a makoto	num-col-size	5	17
	Makako a mašweu a mabedi a makoto	col-num-size	4	13
	Makako a makoto a mabedi a mašweu	size-num-col	4	13
	Makako a makoto a mašweu a mabedi	size-col-num	1	3
	Makako a mašweu a makoto a mabedi	col-size-num	1	3
	(no answer)	*(no answer)*	3	10

'Two thick white loaves'

20	**Meago ye mengwe ye meraro ye metala**	**spec-num-age**	**24**	**80**
	Meago ye mengwe ye metala ye meraro	spec-age-num	3	10
	Meago ye meraro ye mengwe ye metala	num-spec-age	1	3
	Meago ye meraro ye metala ye mengwe	num-age-spec	1	3
	Ye mengwe meago ye meraro ye metala	spec-H-num-age	1	3

'Three other old buildings'

21	**Dikgabo tše nnyane tše mpe tše tala tše disotho**	**size-eval-age-col**	**12**	**40**
	Dikgabjana tše mpe tše disotho	Hsize-eval-col	5	17
	Dikgabo tše disotho tše nnyane tše tala tše mpe	col-size-age-eval	2	7
	Dikgabo tše nnyane tše disotho tše mpe tše tala	size-col-eval-age	1	3
	Dikgabo tše nnyane tše tala tše disotho tše mpe	size-age-col-eval	1	3
	Dikgabo tše tala tše nnyane[2] tše disotho tše mpe	age-size-col-eval	1	3
	Dikgabo tše mpe tše nnyane tše disotho tše tala	eval-size-col-age	1	3
	Dikgabo tše mpe tše tala tše disotho tše nnyane	eval-age-col-size	1	3
	Dikgabjana tše mpe tše tala tše disotho	Hsize-eval-age-col	1	3
	Dikgabo tše mpe tše tala tše disotho	eval-age-col	1	3
	Dikgabo tše tala tše nnyane tše mpe	age-size-eval	1	3
	Dikgabo tše nnyane tše tala tše disotho[3]	size-age-col	1	3
	Dikgabjana tše disotho tše mpe	Hsize-col-eval	1	3
	(no answer)	*(no answer)*	1	3

'Small evil old brown monkeys'

22	**Dihlare tše koto tše telele tše boima**	**width-height-**	**12**	**40**
	Dihlare tše telele tše koto tše boima	height-width-weight	8	27
	Dihlare tše koto tše boima tše telele	width-weight-height	3	10
	Dihlare tše boima tše koto tše telele	weight-width-height	3	10
	Dihlare tše boima tše telele tše koto	weight-height-width	1	3
	(no answer)	*(no answer)*	3	10

'Tall thick heavy trees'

23	**Ditshwene tše nanana tše tona tše (di) befetswego**	**size-gender-rel**	**9**	**30**
	Ditshwenyana tša dipholo/ (di)tona tše (di) befetswego	Hsize-gender-rel	6	20
	Ditshwene tše befetswego tše nanana tše tona	rel-size-gender	6	20
	Ditshwene tše tona tše nanana tše (di) befetswego	gender-size-rel	3	10
	Ditshwene tše befetswego tše tona tše nanana	rel-gender-size	2	7
	Dirotwana tše di befetswego	Hsize+gen-rel	2	7
	Ditshwene tše nanana tše befetswego tše tona	size-rel-gender	1	3
	Dirotwana tše tona tše di befetswego	Hsize-gender-rel	1	3

'Angry young male baboons'

24	**(ke) Ditlou tše kae tše dingwe tše kgolo**	**spec-spec-size**	**7**	**23**
	(ke) Ditlou tše kae tše kgolo tše dingwe	spec-size-spec	6	20
	(ke) Ditlou tše dingwe tše kae tše kgolo	spec-spec-size	4	13
	(ke) Ditlou tše dingwe tše kgolo tše kae	spec-size-spec	4	13
	(ke) Ditlou tše kae tše kgolo	spec-size	3	10
	Ditlou tše kgolo tše kae	size-spec	2	7
	Ditlou tše kgolo tše dingwe (ke) tše kae	size-spec-spec	1	3
	Ditlou tše kgolo tše kae tše dingwe	size-spec-spec	1	3
	Tše dikae tše dingwe ditlou tše dikgolo	spec-spec-H-size	1	3
	Tše dikae ditlou tše dingwe	spec-H-spec	1	3

'How many other big elephants?'

25	**Barwarre ba bantši ba bakaaka**	**specnum-specsize**	**19**	**63**
	Barwarre ba bakaaka	specsize/num	4	13
	Barwarre ba bakaaka ba bantši	specsize-specnum	4	13
	Ba bantši barware ba bakaaka	specnum-H-specsize	1	3
	(No answer)	*(no answer)*	2	7

'Other brothers this big'

26	**Marotho a mannyane a matsotho a boleta**	**size-col-rel**	**10**	**33**
	Marotho a mannyane a boleta a matsotho	size-rel-col	7	23
	Marotho a boleta a mannyane a matsotho	rel-size-col	4	13
	Marothwana a matsotho a boleta	Hsize-col-rel	3	10
	Marotho a matsotho a mannyane a boleta	col-size-rel	2	7
	Marotho a boleta a matsotho a mannyane	rel-col-size	1	3
	Marotho a matsotho a boleta a mannyane	col-rel-size	2	7
	(no answer)	(n/a)	1	3

'Small soft brown loaves'

27	**Dipuku tše tala tše thata bjalo**	**age-eval-adv**	**8**	**27**
	Dipuku tše tala tše thata tše bjalo	age-eval-spec	6	20
	Dipuku tše thata tše tala tše bjalo	eval-age-spec	6	20

	Dipuku tše bjalo tše tala tše thata	spec-age-eval	4	13
	Dipuku tše tala tše thata di bjalo	age-eval-adv	2	7
	Dipuku tše bjalo tše thata tše tala	spec-eval-age	2	7
	Dipuku tše thata tše bjalo tše tala	eval-spec-age	1	3
	Ke dipuku tše tala tše thata	interr-age-eval	1	3

'Such difficult old books'

28	**Ditapola tše nne tše monate**	**num-rel**	**22**	**73**
	Ditapola tše monate tše nne	rel-num	8	27

'Four tasty potatoes'

29	**Barutiši ba bantši ba bohlale**	**spec-rel**	**20**	**67**
	Barutiši ba bohlale ba bantši	rel-spec	10	33

'Manny clever teachers'

30	**Kantoro ye kgolo ya selete**	**size-NR**	**19**	**63**
	Kantoro ya selete ye kgolo	NR-size	9	30
	Kantorokgolo ya selete	Hsize-NR	2	7

'Big regional centres'

31	**Banna ba bašoro ba dipolitiki**	**eval-rel**	**13**	**43**
	Banna ba dipolitiki ba bašoro	rel-eval	12	40
	Boradipolitiki ba bašoro	Nrel-eval	4	13
	(no answer)	(n/a)	1	3

'Cruel political men'

32	**Basemane ba bararo ba Basotho ba borwa**	**num-nat.rel-rel**	**16**	**53**
	Basemane ba bararo ba ba borwa ba Basotho	num-rel-nat.rel	5	17
	Basemane ba borwa ba Basotho babararo	rel-nat.rel-num	3	10
	Basemane ba Basotho ba bararo ba borwa	nat.rel-num-rel	2	6
	Basemane ba Basotho ba borwa ba bararo	nat.rel-rel-num	2	6
	Basemane ba bararo ba Basotho	num-nat.rel	1	3
	Basemane ba Basotho ba bararo	nat.rel-num	1	3

'Three Southern African boys'

33	**(ke) Boramotse ba bakae ba batona ba ma Afrika**	**spec-gender-nat.rel**	**9**	**30**
	(ke) Boramotse ba ma Afrika ba batona[4] ba bakae	nat.rel-gender-spec	6	20

	Boramotse ba batona ba ma Afrika ba bakae	gender-nat.rel-spec	5	17
	(ke) Boramotse ba bakae ba bathobaso/ma Afrika	spec-nat.rel	4	13
	(ke) Boramotse ba ma Afrika ba bakae	nat.rel-spec	2	7
	Ma Afrika Boramotse ba batona ba bakae	natrel-H-gend-spec	2	7
	Boramotse ba batona ba bakae ba ma Afrika	gender-spec-nat.rel	1	3
	(no answer)	(n/a)	1	3

'How many male African mayors?'

34	**Baprista ba batala ba go rapela**	**age-VR**	**21**	**70**
	Baprista ba go rapela ba batala	VR-age	8	27
	(no answer)	(n/a)	1	3

'Old praying priests'

35	**Dinonyana tše sorolwana tša go opela**	**col-VR**	**16**	**53**
	Dinonyana tša go opela tše sorolwana	VR-col	14	47

'Yellow singing birds'

36	**Nku ye nngwe ye tshese ya go ja**	**spec-size-VR**	**21**	**70**
	Nku ye tshese ye nngwe ya go ja	size-spec-VR	4	13
	Nku ye nngwe ya go ja ye tshese	spec-VR-size	3	10
	Nku ya go ja ye nngwe ye tshese	VR-spec-size	1	3
	Ye nngwe nku ye tshese ya go ja	spec-H-size-VR	1	3

'Another thin grazing sheep'

37	**Dikolobe tše tlhano tša go befa tša go kitima**	**num-VRadj-Vrpart**	**18**	**60**
	Dikolobe tša go befa tše tlhano tša go kitima	Vradj-num-Vrpart	5	17
	Dikolobe tša go kitima tše tlhano tša go befa	Vrpart-num-Vradj	3	10
	Dikolobe tše tlhano tša go kitima tša go befa	num-Vrpart-Vradj	2	7
	Dikolobe tša go kitima tša go befa tše tlhano	VRpart-VRadj-num	1	3
	Dikolobe tše tlhano tša go kitima	num-VRpart	1	3

'Five ugly running pigs'

38	**Basetsana ba batelele ba mafolofolo**	**size-rel**	**8**	**27**
	Basetsana ba batelele ba baswa ba mafolofolo	size-age-rel	5	17
	Basetsana ba batelele ba mafolofolo ba baswa	size-rel-age	5	17
	Basetsana ba baswa ba mafolofolo ba batelele	age-rel-size	5	17

	Basetsana ba baswa ba batelele ba mafolofolo	age-size-rel	4	13
	Basetsana ba mafolofolo ba baswa ba batelele	rel-age-size	2	7
	Basetsana ba baswa ba mafolofolo	age-rel	1	3

'Tall energetic young girls'

39	**Sebjanatsopa se sebjalo se segolo se setala sa go kganya sa Polokwane**	**spec-size-col-VR-NR**	**6**	**20**
	Sebjanatsopa se setala se sebjalo sa go kganya sa Polokwane se segolo	col-spec-VR-NR-size	2	7
	Sebjanatsopa se sebjalo sa go kganya se setala se segolo sa Polokwane	spec-VR-col-size-NR	2	7
	Sebjanatsopa se segolo sa Polokwane se setala sa go kganya	size-NR-col-VR	2	7
	Sebjanatsopa se setala sa go kganya se segolo sa Polokwane	col-VR-size-NR	2	7
	Sebjanatsopa se sebjalo se segolo sa go kganya se setala sa Polokwane	spec-size-VR-col-NR	1	3
	Sebjanatsopa se sebjalo se setala sa go kganya sa Polokwane se segolo	spec-col-VR-NR-size	1	3
	Sebjanatsopa se setala se segolo sa go kganya se sebjalo sa Polokwane	col-size-VR-spec-NR	1	3
	Sebjanatsopa se setala sa Polokwane se segolo sa go kganya se sebjalo	col-NR-size-VR-spec	1	3
	Sebjanatsopa se segolo sa go kganya se setala sa Polokwane se sebjalo	size-VR-col-NR-spec	1	3
	Sebjanatsopa sa Polokwane se setala se segolo sa go kganya se sebjalo	NR-col-size-VR-spec	1	3
	Sebjanatsopa sa go kganya se setala sa Polokwane se sebjalo	VR-col-NR-spec	1	3
	Sebjanatsopa se setala se sebjalo sa go kganya sa Polokwane	col-spec-VR-NR	1	3
	Sebjanatsopa se sebjalo se segolo sa go kganya sa Polokwane	spec-size-VR-NR	1	3
	Sebjanatsopa sa Polokwane se segolo sa go kganya se sebjalo	NR-size-VR-spec	1	3

	Sebjanatsopa sa Polokwane se segolo se setala se sebjalo	NR-size-col-spec	1	3
	Sebjanatsopa sa Polokwane sa go kganya se setala se segolo	NR-VR-col-size	1	3
	Sebjanatsopa se segolo se sebjalo sa go kganya sa Polokwane	size-spec-VR-NR	1	3
	Sebjanatsopa se setala sa Polokwane sa go kganya	col-NR-VR	1	3
	Sebjanatsopa se setala sa go kganya sa Polokwane	col-VR-NR	1	3
	(no answer)	(n/a)	1	3

'Such, blue, shining, Polokwane vases'

40	**Banna ba babedi ba bašoro ba maatla ba go hlabana**	**num-eval-rel-VR**	**12**	**40**
	Banna ba babedi ba bašoro ba maatla ba bahlabani	num-eval-rel-VR	3	10
	Banna ba babedi ba maatla ba bašoro ba bahlabani	num-rel-eval-VR	3	10
	Banna ba babedi ba go hlabana ba bašoro ba maatla	num-VR-eval-rel	1	3
	Banna ba babedi ba go hlabana ba maatla ba bašoro	num-VR-rel-eval	1	3
	Banna ba babedi ba maatla ba go hlabana ba bašoro	num-rel-VR-eval	1	3
	Banna ba babedi ba bašoro ba go hlabana ba maatla	num-eval-VR-rel	1	3
	Banna ba bašoro ba maatla ba babedi ba go hlabana	eval-rel-num-VR	1	3
	Banna ba bašoro ba babedi ba maatla ba go hlabana	eval-num-rel-VR	1	3
	Banna ba maatla ba babedi ba bašoro ba go hlabana	rel-num-eval-VR	1	3
	Banna ba go hlabana ba bašoro ba maatla ba babedi	VR-eval-rel-num	1	3
	Banna ba bašoro ba maatla ba go hlabana ba babedi	eval-rel-VR-num	1	3
	Banna ba babedi ba maatla ba bašoro	num-rel-eval	1	3
	Banna ba babedi ba dinatla ba go hlabana	num-rel/eval-VR	1	3
	Banna ba dinatla ba babedi ba go hlabana	rel/eval-num-VR	1	3

'Two strong, cruel, fighting men'

Effects of task repetition

on written language production

in Task Based Language Teaching

Bimali Indrarathne
Lancaster University, UK

Abstract

Different types of communicative tasks are used to promote language learning in Task Based Language Teaching. Repetition of such tasks in class is considered to help learners to reflect on their own language production and thus assists them to improve their performance. It is believed that learners are able to store information related to conceptualisation, formulation and articulation when a task is performed for the first time and this information can be utilised productively when the same task is performed for the second time. Consequently, on the second performance, learners are left with more time to pay attention to other aspects such as fluency, accuracy and complexity of their language production. Several empirical studies on oral task repetition have given positive evidence of increased fluency, accuracy and complexity of task repetition. However, there has been limited research on the impact of repetition on written language production. This paper discusses the results of a case study of written narrative task repetition in which the participant displayed increased performance in accuracy, fluency and complexity of her written language production, in particular in accuracy. The study also reveals that learners are likely to transfer their knowledge of discourse features related to a task when it is performed repeatedly.

Keywords: *Accuracy, fluency, complexity, written task repetition*

1. Introduction

Task Based Language Teaching (TBLT) emphasises the use of different types of communicative tasks in order to promote language learning. Task repetition has drawn much attention as an important aspect of TBLT since it is considered primarily useful in making learners alter their language production (Bygate, 1996; Bygate, 2001; Bygate & Samuda, 2005). Bygate and Samuda (2005) define task repetition as "repetition of the same or slightly altered tasks – whether whole task or parts of the task" (p.43). Bygate (2001) further identifies real task repetition as "the kind experienced by learners when they find themselves repeatedly in highly similar communication situations and with the opportunity to build on their previous attempt at completing the task" (p.29).

Several empirical studies on oral task repetition have investigated its impact on language performance and identified its ability to improve oral language production. One of them is Bygate's (1996) study that investigated the effects of narrative task repetition. On the first occasion, the participant in this study had to watch a short video clip and narrate the story. On the second occasion, she had to watch the clip again and narrate the story again. Bygate (ibid) reports that both fluency and complexity of language performance increased in the second performance; however, the increase in accuracy was not so striking. In a similar study which ran for a longer period with more participants, Bygate (2001) reports similar findings.

Although there have been several empirical studies on oral task repetition, the research on the effects of task repetition on written language performance is limited. This paper discusses the results of a case study that investigated the written language performance of an L2 learner of English who had to perform a written narrative task based on two picture

stories. On the first occasion, the participant wrote a story based on a set of pictures and on the next day, she had to repeat the same task. On day three she wrote another story based on another set of pictures and repeated the task on day four. The participant was not given any input, feedback, or explicit instructions before the first performance or between performances. The performance was measured in terms of accuracy, fluency and complexity of the language produced on each occasion.

2. Literature Review

Levelt (1989) notes that there are four processes in speech production. The first, conceptualisation, is how the message is formed based on the speaker's knowledge of the topic, background, discourse and experience. The next process is formulation i.e. selection of appropriate words, expressions, sentences and even pronunciation to express the message. The third is articulation, which is the use of speech organs such as the tongue, lips, teeth and palate to produce the message. The last is self-monitoring, i.e. speakers are able to monitor their own speech and do self-correction. Bygate (2001, p.28) argues that when a task is performed for the first time, learners are able to store information related to conceptualisation, formulation and articulation of the task in their long term memory and this information is accessible to them when the task is repeated. Thus, they are left with an opportunity to alter their performance on the second occasion. Moreover, they have more freedom and time to pay attention to more redundant grammatical forms and discourse patterns on the second occasion; therefore, the repeated performance may become more accurate.

Even though task repetition is believed to improve language production in terms of conceptualisation, formulation and articulation, Bygate and Samuda (2005) note that the effect of repetition on articulation is generally likely to be minimal because it involves more

automated performance. However, they agree that conceptualization and formulation might be significantly affected by repetition. For example, learners might be able to use the information related to the topic and content stored in their long term memory when conceptualising the task when they perform it for the second time. Furthermore, they might be able to notice more information in the input provided to them when they pay attention to it for the second time. As a result, the formulation process may become speedier leaving more time for learners to self-monitor. For example, in his study of an oral narrative task repetition Bygate (1996) has identified that the participant of the study demonstrated more frequent self correction repetitions of words in the repeated performance. Bygate further points out that this result was likely due to the participant spending less time on content planning and thus having more time to pay attention to word choice and grammatical features on the second performance.

Bygate (2001) also stresses that L2 learners have to primarily create form-meaning relations when producing language and for that they need to draw appropriate morpholexical items from memory, match them with the message that they have to produce and adapt the items if necessary. With this, they need to pay attention to irregularities of natural language and also redundant forms. If learners are familiar with the topic, then the time that they have to spend on conceptualization decreases allowing them to focus more on redundant language forms. As a result, if a task is repeated language production can be improved by means of fluency, accuracy and/or complexity. Fluency, according to Skehan (1998) is primarily related to learners' ability to communicate meaningfully in real time i.e. with minimal hesitations or pauses. Accuracy is how well a learner is able to produce language according to the rules of the target language and complexity relates to how advanced the language produced is, i.e. whether the learner is able to use a range of structures including more redundant forms (Skehan, 1998).

Several empirical studies have investigated the effects of task repetition on accuracy, fluency and complexity of oral language production. Bygate (1996) reports that fluency, complexity and accuracy increased in an oral narrative task repetition; however, the increase in accuracy was non-significant. Similarly, Bygate (2001) in another oral narrative task repetition study has identified a significant improvement in fluency and complexity; but not in accuracy. Birjandi and Ahangori (2008) have used three types of oral tasks: a personal narrative, a story narrative, a decision-making task in a study that involved participants' repetition of tasks and have observed an increase in fluency and complexity, but not in accuracy. In contrast, Matsumara, Kawamura and Affricano (2008) do not report any significant gain in fluency in two types of repeated tasks: a narrative and a decision-making task that they used in their study. However they have noticed improvement in accuracy and complexity in both types of tasks with a significant improvement in accuracy in the narrative task and a significant improvement in complexity in the decision-making task. Hawkes (2011) has identified that when a form-focused session is included in between the first and second performance of oral tasks, the participants are likely to focus more on accuracy in the second performance. Lynch and Mclean (2000) who believe that the intervals between the first and second performances of the tasks have an effect on production, have identified that the immediate repetition of a task could increase accuracy, performance in pronunciation, self correction and vocabulary selection. In summary, different empirical studies have provided evidence on the effects of task repetition on accuracy, fluency and complexity of oral language production.

A few studies have also investigated the impact of task repetition on written language production. For example, Jung (2013) reports a study that was focused on written language production through repetition of essays. This study highlights the fact that that task repetition was not able to increase accuracy of written language production, but was able to increase

fluency and complexity of it. Larsen-Freeman (2006) notes that the participants in her study, as a group, demonstrated that accuracy, fluency and complexity increased when a narrative task was repeated, first in the written mode and then in the oral mode. Since there is less empirical research on written language production, it is difficult to clearly hypothesise the extent to which task repetition can increase accuracy, fluency and/or complexity of language production.

Relating to assumptions made by psycholinguists, conversation analysts, ethnomethodologists and SLA theorists, Bygate and Samuda (2005) also suggest that when a particular type of communication task is repeated, information on discourse features such as narrative structures which are also stored in the long term memory store are likely to be accessible to the learners. Bygate and Samuda's (2005) and Bygate's (1996) studies have indicated an improvement in how learners use discourse features when the same task is repeated. For example, learners were able to produce better stories in terms of the use of discourse features in the repeated performance when narrative tasks were repeated. For example, Bygate and Samuda (2005) paid attention to how discourse complexity (e.g. discourse features such as evaluation, interpretation, summarizing and cohesive links) was affected by repetition and identified a significant increase of performance in how learners frame information when the task was repeated. Bygate (1996) also noticed a significant increase in evaluative comments and the use of cohesive devises by the participant in the second performance of the narrative task that he used in his study. Thus, it is evident that learners are able to utilise the discourse features that they learnt/practised in the first performance when they perform the same task on another occasion.

Several empirical studies on task repetition indicate that the type of task that is repeated also has a significant impact on performance. For example, Matsumara, Kawamura and Affricano (2008), in their study of Japanese EFL learners, state that a narrative task

repetition could reduce lack of fluency of the participants; however, a decision-making task repetition could not do so. In contrast, the decision making task could increase the complexity of language to a greater extent the narrative task. Birjandi and Ahangori (2008), in their study of a personal narrative, a story narrative and a decision making task repetition by Iranian EFL learners have also identified that the personal narrative task was able to increase accuracy and complexity more than the other two types and that the task type did not make any significant influence on fluency. However, their use of personal task (in which learners were asked to explain a personal activity such as asking someone to turn off the oven in the flat where they live) may have an effect on the results since topic familiarity can affect performance in such tasks. Bygate (2001) in his study has identified that practising a particular task type does not have any significant effect on language production. Gass and Mackey (1999) have also reported less impact of task type repetition. For example, Gass and Mackey (1999) have identified that the participants in their study of narrative repetitions could increase their performance in holistic judgement and morphosyntax (use of *estar* in Spanish) when the same tasks were repeated, but did not transfer their ability to a similar type of task.

Taking the literature discussed in this paper into account, it is possible to state that most empirical studies have identified significant effects of task repetition on fluency, accuracy and complexity of oral language production. The learners have further demonstrated their ability to utilise discourse features that they learnt/used in the first performance when they repeat the same task. The studies discussed here further suggest that the type of task that is repeated also has an impact on the performance. For example, it is evident that narrative tasks provide more evidence on the positive effects of task repetition. Although the effect of task repetition on oral language production has been investigated, little research to date has investigated its effect on written language production. Taking this into account, a case study

was designed to analyse the effects of narrative task repetition on written language production.

Research questions

- Does task repetition increase fluency of written language performance?

- Does task repetition increase accuracy of written language performance?

- Does task repetition increase complexity of written language performance?

3. Methodology

3.1 Participant

Mojavesi (2013) reports that learners with higher L2 proficiency demonstrated more improvement in accuracy, fluency and complexity compared to the lower L2 proficiency level learners in an oral task repetition study. Jung (2013) used two groups who engaged in task repetition i.e. a group who received feedback between sessions and a group who did not receive feedback. One of the two participants in the second group was in the higher intermediate level and the other in the lower intermediate level. The participants in Larsen-Freeman's (2006) study were also in the higher intermediate level while Matsumara, Kawamura and Affricano (2008) used advanced beginners in their study. Considering the fact that most of these empirical studies have used either higher proficiency level learners or lower proficiency level learners, the present study was focused on a learner at a mid proficiency level.

The participant of this study was a 26 year old Greek female who was reading for a master's degree in management at a UK university. Her IELTS score was 6.5 and she did not follow any English language course during the period of this study. She had learned English

as a foreign language for nearly 16 years in Greece, but had not learned any other foreign language. By the time the study was conducted, she had lived in the UK for five months.

3.2 Procedure

Two picture stories were used in the study and both stories contained six pictures each. Story A (Appendix A) was used on the first two days and story B (Appendix B) was used on day three and day four. The participant had to write a story based on the pictures in A on day one, repeat it on day two and write a story on the pictures in B on day three and repeat it on day four. A pre or a post-test was not conducted in the study and the analysis was purely done based on the four written stories produced by the participant. On all occasions, the participant was not allowed to see the previous writings; however, she could refer to the pictures while writing. Furthermore, no linguistic input, feedback, or instructions were given to the participant before the first performance or between performances.

On the first day, the participant was told that she needed to write this story to be sent to a children's magazine and on the second day, she was told that she needed to rewrite it. However, she was not allowed to see the story that she wrote on the previous day. On the third day, she was asked to write another story to be sent to a primary school teacher who could use it in her class and similar to the procedure on day two, on the fourth day, the participant was asked to rewrite the story. The participant was asked to write a 'story for a children's magazine' and 'to be used in a primary class' in order to provide a purpose for her writing.

The interval between each task repetition was approximately 24 hours. On all four days, the participant was informed that the maximum time available for the task was 45 minutes. The participant was not reminded about the time remaining while she was writing. The study was conducted in the same room with only the participant and the researcher

present, and without any disturbances. The participant was not allowed to use any other resources such as dictionaries, books or the internet when writing. Thus, the tasks were repeated under the same conditions. After all four phases of repetition, the participant was interviewed to obtain her views on her experience and the answers are analysed in the discussion section of this paper.

3.3 Instruments

Narrative tasks (picture stories) were used as the instrument of this study for two reasons: firstly it is the most common type of task that has been used in empirical studies on oral task repetition, and secondly most of these studies highlight that repeating narrative tasks have an impact on language performance (Bygate, 1996; Bygate, 2001; Bygate & Samuda, 2005; Matsumara, Kawamura & Affricano, 2008; Birjandi & Ahangori, 2008; Gass & Mackey, 1999). Moreover, Bygate (as stated in Matsumara, Kawamura & Affricano, 2008, p.130) states that narrative tasks "invite linguistically denser talk" contributing to L2 development. Furthermore, Kawauchi (as stated in Matsumara, Kawamura & Affricano, 2008, p.130) points out that narrative tasks can minimise individual variations in language production.

It was also important to select a task which has the qualities of tasks used in TBLT for this study. A task is defined by Skehan (1998, p.95) as "an activity in which: meaning is primary; there is some communication problem to solve; there is some sort of relationship to comparable real world activities; task completion has some priority; and the assessment is in terms of task outcome." In this study, the participant was asked to write two stories based on series of pictures to be published in a children's magazine and to be used in a language class. Therefore, the writer had to pay attention to meaning which is the primary focus of the task. Since she had to describe the pictures by analysing the underlying story and also had to make

links between the different scenes, there was a communication demand for the participant. Furthermore, the task resembles real world activities because story telling takes place in real world contexts mainly in the form of personal narratives. Task completion was given priority by asking the participant to complete writing the story for a purpose i.e. the task should have been completed to serve the purpose (sending it to a magazine, sending it to a teacher). The analysis of language proficiency of task repetition was done based on the outcome (the written stories) of the task. Thus, the narrative tasks used in the study can be considered compatible with the definition offered by Skehan on 'a task.'

The results obtained in the task repetition phases were analysed quantitively in order to answer the research questions. The other instrument of the study was the interview conducted with the participant and the answers were qualitatively analysed to investigate the participant's experience of task repetition.

3.4 Data analysis methods

Data analysis of the study was done based on fluency, accuracy and complexity of the written stories produced by the participant. Housen and Kuiken (2009) mention that fluency, accuracy and complexity measurements are the major research variables used in applied linguistic research in order to measure learners' proficiency and progress in language learning. Thus, those measurements were used to analyse the written language production of the participant of the present study.

Several researchers have used different measures to analyse fluency, accuracy and complexity of written language (Ellis & Yuan, 2004; Larsen-Freeman, 2006; Storch & Wigglesworth 2007). The main measurement tools (which are also called 'production units') used in such analyses are T units and independent and dependent clauses (Wolfe-Quintero,

Inagaki & Kim, 1998). A T-unit is defined by Hunt (as stated in Bardovi-Harlig, 1992, p.390) "as a combination of an independent clause and its dependent clauses." An independent clause, according to Richards, Platt and Platt (as stated in Storch & Wigglesworth, 2007) is a clause that can stand on its own and a dependent clause is a clause which has to be used with another clause to form a grammatical sentence (ibid). Storch and Wigglesworth (2007, p.160) state that an independent clause minimally consists of a finite verb and a dependent clause minimally consists of a finite or non-finite verb element and "one other clause element (subject, object, complement or adverbial)." Based on these definitions, T units and clauses were identified in the written scripts in this study.

3.4.1 Fluency measurements

Fluency in written language appears to be slightly different from fluency in oral language (see the literature review for the explanation of oral fluency). For example, Wolfe-Quintero, Inagaki and Kim (1998) state that fluency in writing can be measured through number, length or rate of the production units in the text. As Larsen-Freeman (as stated in Wolfe-Quintero, Inagaki & Kim, 1998, p.14) mentions, a fluent writer is considered to have the ability to write longer compositions; thus, length of the composition is a measure of fluency. Therefore, one length measure (number of words per T unit) of fluency was used in this study. However, time factor plays an important role in length measurements because the length depends on whether the learner used the maximum available time or otherwise. In the present study, even though the participant was given a time limit, the nature of the task made the learner use less time to complete the composition when the tasks were repeated. Therefore, a ratio fluency measure was also necessary for the analysis. Hence, number of words per minute which is considered to have a strong positive correlation with proficiency

was also used for the analysis (This was first proposed by Arthur as a fluency ratio measure (as stated in Wolfe-Quintero, Inagaki & Kim, 1998, p.14)).

3.4.2 Accuracy measurements

In order to measure text accuracy, the primary method used in empirical studies is calculating the number of errors in a composition. As an accuracy frequency measure, the number of error-free T units was used in this study. This measurement is proposed by Wolfe-Quintero, Inagaki and Kim (1998) as a development measure that has a strong positive correlation with proficiency. Ratio measures such as error-free T units per T unit and error-free clauses per clause are also frequently used to measure accuracy; however, they are considered to be suitable for analysing accuracy in long term studies (ibid). Since the present study is a short term study aimed at measuring the overall accuracy of texts, errors per T unit was used as a ratio measure because it correlates with holistic ratings (ibid). As proposed by Storch and Wigglesworth (2007), syntactic (word order, missing words) and morphological (verb tense, subject-verb agreement, articles, prepositions and word forms) errors were identified as errors for calculations.

3.4.3 Complexity measurements

Storch and Wigglesworth (2007) emphasise the importance of analysing the complexity of language production when measuring proficiency because of the possibility of learners refraining from using complex structures to achieve higher accuracy. Thus, complexity measurements can determine whether the learner is willing to experiment with complex forms of language. Wolfe-Quintero, Inagaki and Kim (1998) state that there are two types of complexity measure: grammatical and lexical. Proportion of clauses to T unit (grammatical complexity) and lexical sophistication (lexical complexity) proposed by them

as reliable complexity measures were used in this study. It was necessary to judge whether the words were sophisticated or not in order to measure lexical sophistication. Therefore, Oxford 3000 words list which includes the most frequent English words (Turnbull, 2010) was used to determine sophistication of words. The words which belong to this list were not considered sophisticated. In summary, data analysis was done based on the measurements given in Table 1.

Table 1

Fluency, accuracy and complexity measures

Fluency	Accuracy	Complexity
number of words per T unit	error-free T units	proposition of clauses to T unit
number of words per minute	errors per T unit	lexical sophistication (total number of sophisticated lexical words/ total number of lexical words)

3.5 Data coding

The researcher coded the data for the first time and a second coder was used to ensure the reliability of coding. The intra-coder reliability for T-unit measurements varied from 94% to 97%. Then a third coder was used and the intra-coder reliability between the first and third coders for T unit measurements was 98% to 100%.

4. Results

The results of the study are summarised under each measurement category. Considering the small amount of data involved in this case study, a statistical analysis was not conducted. It is also worth noting here that the participant spent 35, 25, 20 and 18 minutes

respectively to complete the four performances without the researcher reminding her of the amount of time left.

4.1 Fluency

As can be seen in Table 2 below, the number of words per T unit increased in the second performance compared to the first in both tasks; however, the increase in the second task is less striking. Number of words per minute also indicates a clear increase in the second performance compared to the first in the first task and a slight increase in the second task. The results of the fluency measures of the two tasks demonstrate that repetition of both tasks impacted the participant's written language fluency.

Table 2

Fluency measures

Task	Performance	Number of words per T unit	Number of words per minute
1	1	10.08	6.914
	2	12.26	9.320
2	1	10.40	13.000
	2	10.45	13.940

4.2 Accuracy

The error free T units measurement highlights an increase in the second performance compared to the first in both tasks. The errors per T unit measure demonstrates a decrease in the second performance compared to the first in both tasks. This indicates that task repetition can make a positive impact on accuracy of written language production.

Table 3

Accuracy measures

Task	Performance	Error free T units %	Errors per T unit
1	1	37.50	0.70
	2	57.89	0.42
2	1	72.00	0.36
	2	83.33	0.20

4.3 Complexity

Proportion of clauses to T units demonstrates an increase in the second performance compared to the first in both tasks. Lexical sophistication also demonstrates similar results. However, the increase in lexical sophistication is clear in the second task compared to the first.

Table 4

Complexity measures

Task	Performance	Proposition of clauses to T units	Lexical sophistication
1	1	1.25	0.016
	2	1.47	0.017
2	1	1.28	0.038
	2	1.37	0.055

5. Discussion

It is noteworthy that the findings discussed in this paper are of a case study and thus it is difficult to generalize them without analysing the language production of a large group of participants. However, the study provides some useful insights into the effects of task

repetition on written language production. The results of this study, in summary, indicate that task repetition could increase fluency, accuracy and complexity in written language production. In particular the two accuracy measurements i.e. error free T units and errors per T unit indicate a striking increase in accuracy in both repetitions. Complexity also demonstrates an increase in both tasks. However, the increase of performance in fluency is less striking particularly in the second task. The less striking increase of fluency could be a result of possible 'trade off' effect (Bygate, 2001): the gains in accuracy and complexity must have been paid for by a loss or by a small gain in fluency. For example, the participant of this study spent 20 and 18 minutes respectively in the two performances of task two and the difference between times is not significantly different. Therefore, it is possible to predict that the participant may have paid more attention to accuracy and complexity in the second performance and thus could not demonstrate a striking improvement in fluency.

The results of this study are in contrast with the results of three main oral narrative task repetition studies. For example, Bygate (1996, 2001) and Birjandi and Ahangori (2008) reported that both fluency and complexity of oral language performance increased when oral narrative tasks were repeated in their studies. However, the gains in accuracy were minimal in all these studies. In contrast, the written narrative task repetition (this study) indicates an increase in accuracy and complexity and a less striking increase in fluency. This could be due to the mode of output: in oral task performance, learners pay more attention to speed of articulation and in written task performance, they pay more attention to grammatical accuracy of sentences.

Bygate (2001) and Gass and Mackey (1999) indicate that task type repetition has less impact on language production i.e. the features learnt by performing one task cycle cannot be or may not be transferred when performing a similar task type. However, the present study indicates that there could be an effect of task type repetition on written language production.

For example, the participant's performance in number of words per minute, error free T units, errors per T unit and lexical sophistication demonstrates a clear difference between the second performance of the first task and first performance of the second task. Number of words per minute (fluency), error free T units (accuracy) and lexical sophistication (complexity) increased in the first performance of the second task compared to the second performance of the first task and errors per T unit (accuracy) decreased in the first performance of the second task compared to the second performance of the first task. This could be due to the participant transferring the abilities gained in performing the first task to the performance of the second task. Thus, it is possible to state that task type repetition might also assist learners to improve written language performance.

It is also important to analyse the post interview data in order to understand how the participant viewed the task repetition process. The participant stated that it was easier for her to write the second versions of the stories due to the fact that she needed less time for planning 'what to write.' This resembles Bygate's (1996) finding in which he states that the participant of his study could perform better on the second occasion because he spent less time on content planning in the repeated performance. The quantitative data in the present study also indicate that the participant could perform better in the second performance.

The participant also stated that she could remember what she wrote in the previous task which helped her to 'change certain words and sentences.' This resembles Bygate's (2001) arguments that learners can utilise information stored in long term memory when repeating a task and learners are left with more time to pay attention to redundant forms when a task is repeated. One such occasion of the present study was that the participant used the word 'disappeared' (which is a more suitable word for the situation) in the second performance of the second task to explain that the two children could not see their food in the basket which she did not use on the first occasion to explain the same situation. The

participant further stated that writing stories became easier and easier because 'I know how to do it now.' This is apparent in the narrative structure of the stories that the participant wrote. For example, there is an increased use of connectors such as 'and, when, after few minutes' in the second performances of the stories to connect events in them. This could be due to, as Bygate (2001) also suggests, the possibility of the participant accessing information on discourse features which are also stored in long term memory.

6. Pedagogical implications

This study brings out some useful insights on how task repetition can be applied in second language classes. It clearly demonstrated that the repetition of a written narrative task could increase accuracy, fluency and complexity of written language production to a greater extent. Therefore, it would be useful for language teachers to use such task repetitions in order to increase the written language production of their learners; in particular, to increase accuracy.

The study also indicated that task type repetition might also increase the performance of written language production of L2 learners. Therefore, task type repetition in class could be worth trying in order to improve the written language production of learners. It is also noteworthy that task type repetition might not be able to increase language production in all aspects related to fluency, accuracy and complexity discussed in this paper since this study also indicated an increase of performance in some of these aspects only. Moreover, task type does not imply the same difficulty; therefore, performance could vary depending on the difficulty level of tasks.

It is also important for teachers to bear in mind that all tasks might not give the same results as discussed here. For example, narrative tasks are considered contributing to L2

development minimizing individual variations in language production (Matsumara, Kawamura & Affricano, 2008). Further, narrative tasks have been successfully used in oral task repetition studies. This study indicated that narrative tasks could increase the performance of written language production of the participant as well. Therefore, it is highly likely that narrative tasks can be successfully used in task repetition. However, other types of tasks might not bring the results brought by narrative task repetition. Thus, teachers have to be careful in selecting tasks.

Teachers also have to bear in mind that the type of narratives used for written language production might influence the lexis that students may produce and thus, it is important to analyse the vocabulary that could be elicited in a task especially if the repetition aims at vocabulary development. For example, lexical complexity achieved by the participant in this study is significantly higher in the second task than in the first task. It is likely that the type of incidents, scenes, and objects included in the pictures have made an impact on lexical sophistication in the performance of the two tasks. Furthermore, the participant used spoken language utterances such as 'oh god' and 'bad luck' in the stories which might be worth considering when narrative tasks are utilized to improve the written language production of learners in real class contexts.

This study also indicated that learners might not fully utilise the full amount of time available for completion of the task especially when it is repeated. This was evident in the present study because the participant decreased the time spent in each performance. This could be because the participant was over confident, less interested in repeating the performance, had lost motivation or sense of challenge or had gained maximum ability. If the reason is one or more of the first two, teachers can make sure that learners pay enough attention to using the extra time available for them maximally to increase their performance. Providing some guidance on what language aspects could be improved in the second

performance could be useful in this regard. It is also important to provide a clear purpose for the second performance to make it more meaningful for learners which could also increase their interest. Teachers might also need to raise the stakes increasing the challenge to keep students focused.

7. Conclusion

Several empirical studies conducted based on oral task repetition provided positive results of language development of learners in terms of accuracy, fluency and complexity. However, a trade-off effect was also quite apparent in the findings. Further, there was less evidence on the positive effects of task type repetition. The present study on written narrative task repetition indicated that task repetition may increase the written language production in terms of accuracy, fluency and complexity and furthermore task type repetition may also positively affect the performance of a similar task later.

Thus, it could be useful for language teachers to utilize written task repetition in order to increase written language development of their learners. However, it is also important for them to carefully select the types of tasks that could be repeated. Furthermore, it is better if the teachers can provide a clear purpose for the learners to repeat a task otherwise the repetition process could be less interesting to the learners.

This study also has several limitations. For example, it was limited to one participant; therefore it is difficult to generalize the results to broader contexts. Furthermore, the participant had constant exposure to English while participating in the study since she lived in the UK during the time of the study which might also have affected the results. Moreover, the English language level of the participant may have played a role in her performance; the

results of a similar study with participants from another English language level may bring different results. The study was also not conducted in a real classroom situation so the results might not be valid to real class contexts as well. Thus, it would be useful to conduct studies with more participants in real classroom contexts to examine the effects of narrative task repetitions on written language production.

References

Bardovi-Harlig, K. (1992). A Second Look at T-Unit Analysis: Reconsidering the Sentence. *TESOL Quarterly*, *26*(2), 390-395.

Birjandi P. & Ahangari, S. (2008) Effects of Task Repetition on the Fluency, Complexity and Accuracy of Iranian EFL Learners. *The Asian EFL Journal, 10*(3), 28-52.

Bygate, M. (1996). Effects of task repetition: appraising the developing language of learners. In J. Willis and D. Willis (Eds.) *Challenge and Change in Language Teaching* (pp. 136-146)*.* Oxford: Heinemann.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan & M. Swain. *Task-based learning: language t*eaching, *learning and assessment* (pp. 23-48). London: Longman.

Bygate, M. & Samuda, V. (2005). Integrative planning through the use of task repetition. In R. Ellis (Ed) *Planning and task performance in a second language.* Amsterdam/New York: Benjamins.

Ellis, R. & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *SSLA*, 26, 59–84.

Gass, S. and Mackey, A. (1999). The effect of task repetition on linguistic output. *Language Learning. 49,* (4). 549-581.

Hawkes, M.L. (2011). Using task repetition to direct learner attention and focus on form. *ELT Journal*, *59*, 1-10.

Housen, A. & Kuiken, F. (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics, 30*(4), 461–473.

Jung, S. (2013). The effect of task repetition and corrective feedback in L2 writing: a pilot study. MSU Working Papers in SLS, *4*, 24-38.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590–619.

Levelt, W.J.M. (1989). *Speaking: from intention to articulation*. Cambridge: The MIT Press.

Lynch, T. & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research, 4*, 221-250.

Matsumara, M., Kawamura, K. & Affricano, A. (2008). Narrative task type repetition and changes in second language use in a classroom environment: a case study. *The Educational Sciences, 10,* 124-145.

Mojavesi, A. (2013). The relationship between task repetition and language proficiency, *Applied Research on English Language, 3*(1), 29-40.

Skehan, P. (1998). A cognitive approach to language learning. Oxford: Oxford University.

Storch, N. & Wigglesworth, G. (2006). Writing tasks: the effects of collaboration. In M.D.P. Mayo (Ed.) *Investigating tasks in formal language learning* (pp. 157-177). Clevedon: Multilingual Matters Ltd.

Tavakoli, P. & Foster, P. (2008). Task Design and Second Language Performance: The Effect of Narrative Type on Learner Output. *Language Learning*, *58*(2), 439–473.

Turnbull, J. (Ed.) (2010). *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.

Wolfe-Quintero, K. Inagaki, S. & Kim, H. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy.* Hawaii: University of Hawaii.

Appendix A
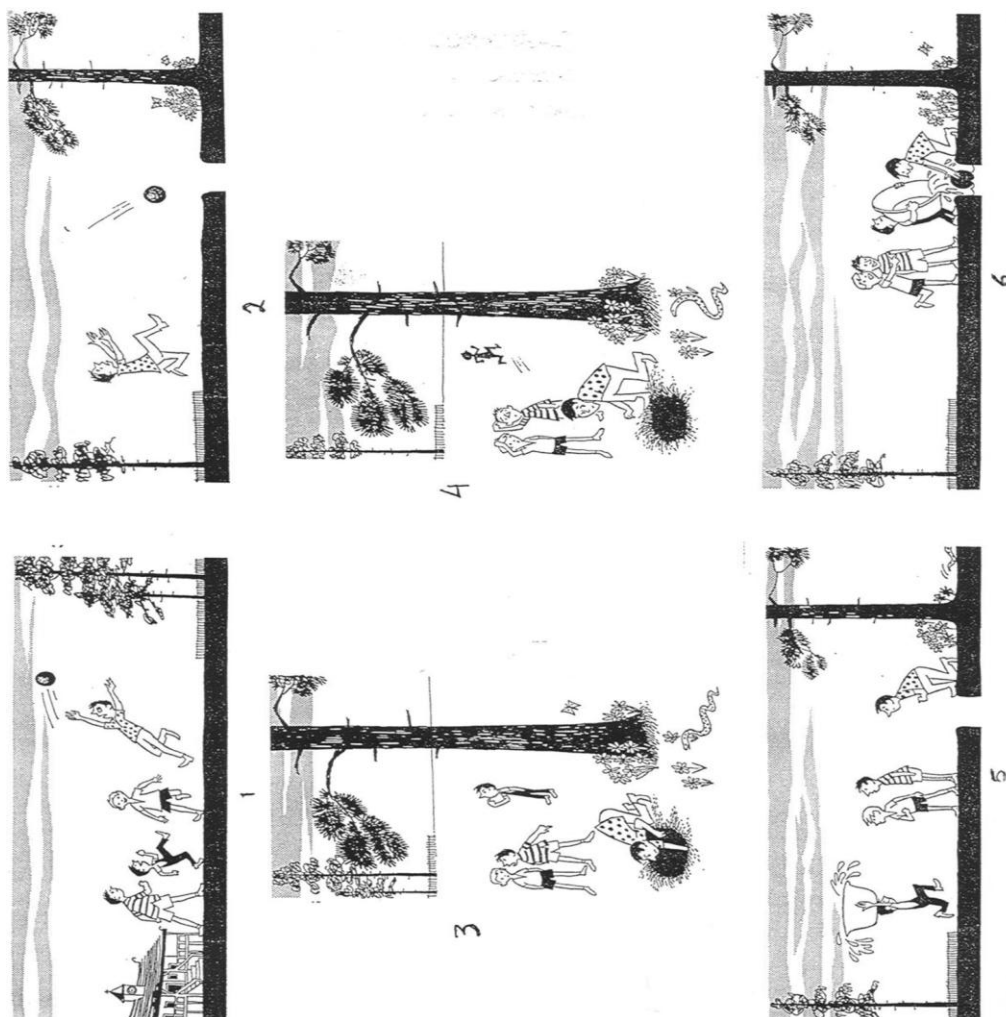
Figure A4 Football Task, Heaton, 1966.

Tavakoli & Foster (2008, p. 473).

Appendix B
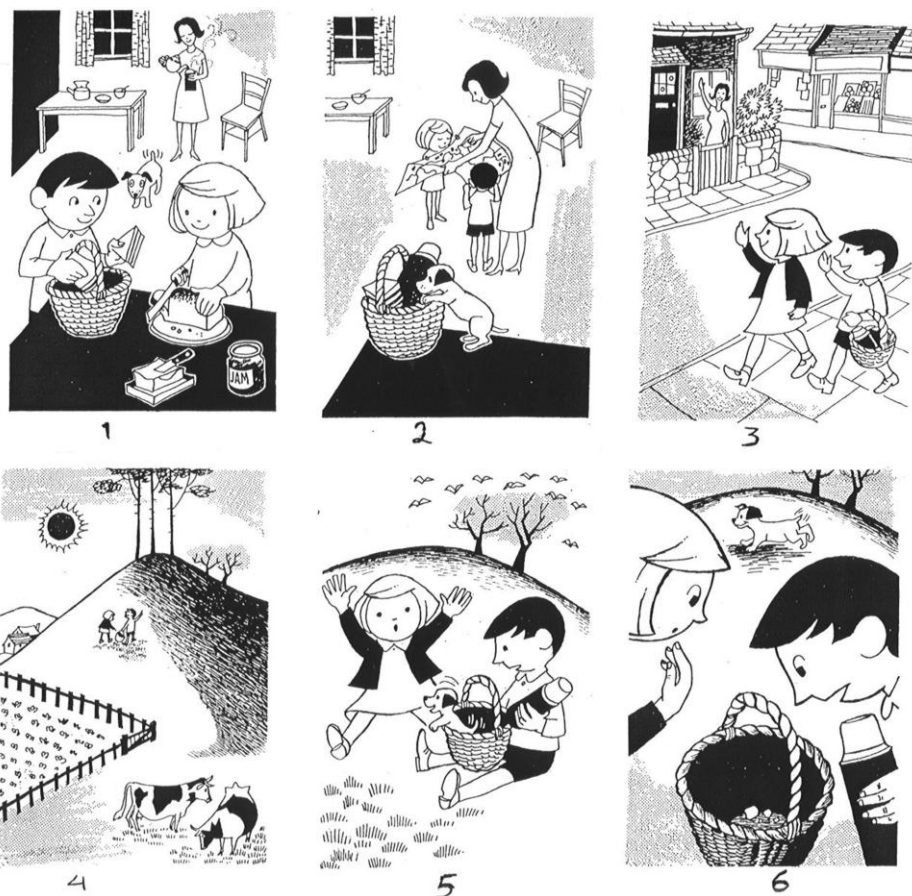
Figure A2 Picnic Task, Heaton, 1966.

Tavakoli & Foster (2008, p.447)

Investigating the listening construct underlying listening-to-summarize tasks

Anchana Rukthong

Lancaster University, UK

Abstract

A major concern in language assessment is that test results or scores may not be generalizable to real-world language use, in view of the fact that test situations may be inherently different from more authentic settings. If this were the case, tests would probably yield invalid results. In this regard, this pilot study aims to shed light on the construct underlying listening-to-summarize tasks and the extent to which this task type can capture the processes used in real-world communication. Four Thai ESL learners participated in the study. They were asked to perform four listening-to-summarize tasks: two requiring an oral summary of listening and two requiring a written summary. Immediately after each task's completion, stimulated recall was conducted. The results show that the participants engaged in eight processes in their attempts to complete the tasks. These processes can be categorized into three main categories of processing: linguistic, semantic and discourse processing. The paper concludes that listening-to-summarize tasks can tap into the processes which are utilised by proficient listeners and necessary for successful academic listening.

1. Introduction

Integrated test tasks, which include both receptive and productive language skills in task performance, are said to mirror academic literacy activity and reflect the ability to use language in authentic situations (Brown, Iwashita, & McNamara, 2005; Lewkowicz, 1997; Plakans, 2008; Plakans & Gebril, 2012; Weigle & Parker, 2012; Weir, 1990). Describing the construct underlying integrated test tasks is, however, not straightforward. Since at least two language skills are involved in task performance, it remains ambiguous what abilities are truly assessed by this task type and what abilities contribute to either success or failure in performance. In fact, when the underlying construct is not clearly understood or well defined, it is difficult for test developers to support their claims about construct representation and relevance and the usefulness of their tests. In this regard, the present study investigates the construct or abilities assessed by this task type by looking into the cognitive processes and sources of knowledge employed to complete the tasks.

To justify the meaning and value of test scores, it is crucial to study construct validity. A fundamental feature of construct validity, as discussed by Messick (1995), is construct representation, which might not be achieved only through relevant content and operative processes, such as examining the correlation of test scores with other external measures. A set of construct indicators, including cognitive processes, strategies and knowledge (metacognitive or self-knowledge) that are applied in task performance, is needed to explain the construct underlying test tasks (Messick, 1995). Thus construct validation has to take into account the meaning of test scores, not only in relation to test items but also to test takers and the context of assessment (Messick, 1995). This is needed in order to provide evidence and a rationale to support the trustworthiness of score interpretation and use (Messick, 1995). By investigating construct validity in this manner, research can explain the degree to which

interpretations and inferences made on the basis of test scores are appropriate and plausible (Bachman & Palmer, 1996; Chapelle, 1999; Messick, 1995; Xi, 2008). In addition, it can point to evidence that supports or discounts inferences or arguments made on the basis of test results (Messick, 1995).

The present study, in particular, focuses on the listening construct underlying listening-to-summarize tasks that include academic lectures as input. Listening is required in a variety of communicative events in academic settings, e.g. lectures, group discussions, tutorials, seminars and meetings with a supervisor. If students are to participate successfully in academic communication, they must have the ability to process and respond to spoken language (Lynch, 2011). Despite its importance, listening remains the least understood of the four language skills because of its ephemeral nature which is not directly observable (Buck, 2001; Field, 2013; Lynch, 2011; Rost, 2002). Though previous research has attempted to identify the construct underlying integrated test tasks (e.g., Brown et al., 2005; Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Frost, Elder, & Wigglesworth, 2011; Gebril, 2010; Gebril & Plakans, 2013, 2014; Plakans & Gebril, 2012), only a few studies aimed to investigate the construct of test tasks integrating a listening source text (e.g. Brown et al., 2005; Cumming et al., 2004; Frost et al., 2011). As for the studies that focus on integrated listening, they rely mainly on linguistic analysis of task performance to identify the test construct. None of them appeared to investigate participants' mental processes using stimulated recall. As acknowledged by language test educators (e.g. Bachman & Palmer, 2010; Messick, 1995), an investigation of cognitive processes utilized by the test takers is necessarily important in the description of the construct underlying the task due to the fact that it reveals thinking processes and knowledge used to complete the test tasks. Further research along this line is thus warranted.

Informed by Messick's (1995) construct validity, the study conceptualizes the construct underlying the listening-to-summarise tasks as the mental processes that test-takers engage in while performing the tasks. Data on test takers' cognitive processes is gathered and described using a cognitive process framework for listening. Literature related to the cognitive process framework and contributing factors of effective L2 listening is reviewed in the next section. Information concerning the research instruments, participants, data collection procedures and analysis is included in the research methodology section, followed by a discussion of research findings and a conclusion.

2. Literature Review

2.1 Cognitive process framework for listening

The description of the cognitive processes involved in listening-to-summarize tasks in the present study is framed by Field's (2013) cognitive processing framework. This model has been adopted for the following reasons. Firstly, it is a listening-based framework that takes into account both individual traits and the interaction between a listener and listening task, which is crucial when describing the listening construct (Buck & Tatsuoka, 1998; Rost, 2002). Secondly, this model has been established and modified on the basis of the processes used by proficient listeners in various contexts, including an academic environment, which is the context that the present study aims to generalize. Lastly, the model notes the role of higher-level processes which are required in real-world academic listening and which the tasks employed in the present study aim to tap into.

In this cognitive processing framework (see Figure 1), Field (2013) explains that successful listening performance entails five main levels of processing: 1) input decoding, 2) lexical search, 3) parsing, 4) meaning construction, and 5) discourse construction. These five

levels, as presented in the shaded boxes, are subdivided into lower-level processes and higher-level processes.
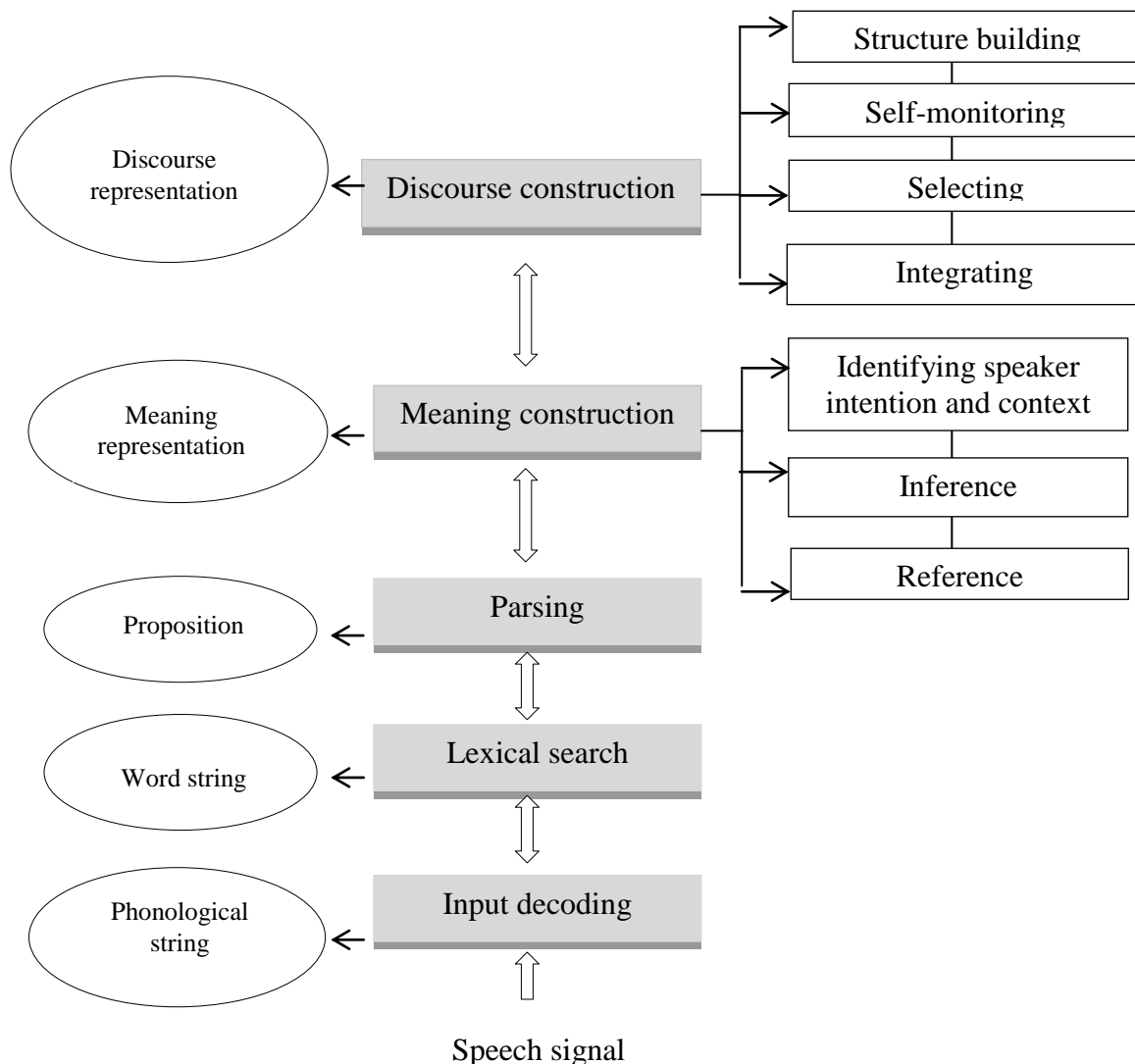
Figure 1: Field's cognitive processing framework for listening adapted from Field (2013)

Lower-level listening processes or linguistic processing involves the first three levels from the bottom (input decoding, word search and parsing), occurring when a message is being decoded into language. Higher-level processes, the top two processes, are associated with meaning and discourse construction. Although the processes are presented in a linear

order, it does not necessarily mean one stage of processing waits for one or more others. Language processes, as noted in this model, often act in a parallel and interactive manner. The numbering is thus used only to represent the levels. The oval shaped boxes in the figure indicate the output of each stage of processing.

Lower-level processes

Lower-level processes, according to Field (2013), involve three levels of linguistic processing, input decoding, word search, and parsing. Field (2013) indicates that listening processing starts from recognizing acoustic input and developing this to obtain a phonological string via input decoding, a set of words from lexical searching, and an abstract proposition via parsing. In input decoding, proficient listeners depend on their phonological knowledge to access a sequence of speech-like sounds and convert these sounds into representations that match the phonological system of the language being spoken (Field, 2013). At this level of processing, the listeners recognize a string of phonemes, some of which are marked as syllables of words. In a lexical search, the listeners map sounds to spoken word forms. Based on their lexical knowledge, the listeners have to determine word boundaries and identify words which are either content or function words in connected speech. At the level of parsing, the listeners segment units in the connected speech and construct propositions by applying their syntactic knowledge, understanding of standard word order, and intonation group boundaries.

Higher-level processes

Higher-level processes involve two levels of processing, meaning and discourse construction (Field, 2013). Listeners start to construct the meaning of what they have heard by relating the propositions obtained from lower-level processing, which is context-

independent, to their own schemata or concepts of knowledge they have developed. At this level, it is the task of the listeners to relate the propositions to the circumstances in which they were produced in order to extract their full meaning and relevance. The raw meaning of the speaker's words is often insufficient to convey the complete meaning of a text (Field, 2013). The listeners, therefore, have to supply additional information to comprehend what is said in a number of ways. One way to do this is to use pragmatic forms of language to interpret the speaker's intention. The listeners may also have to use contextual and semantic knowledge to relate propositions to the context in which they occur. The listener may, in addition, have to infer what the speaker left unsaid from what they have just heard or backtrack to what was being said or what was said earlier.

Discourse construction relates to four processes that the listeners apply to construct their understanding of a spoken text. As presented by Kintsch and van Dijk (1978), these processes are selecting, integrating, self-monitoring and structure building. Selecting is when the listeners assesse the relevance of an incoming piece of information, e.g. whether it is the repetition of a point made earlier or the central point of the topic being developed. On the basis of this consideration, the listeners may store the information being processed or discard it as irrelevant. Integrating is when the listener adds one or more new pieces of information to the discourse representation being developed. It involves recognizing conceptual links between incoming information and that already processed. Self-monitoring entails comparing whether a new piece of information is consistent with what has been processed before. If not, the listener has to consider whether the new judgement is correct or question whether what they have understood earlier and recall is correct. Structure building is when the listeners have to prioritize and organize the information they have stored according to its importance and relevance. A more proficient listener is able to build a more complex information structure than a less proficient one.

Field (2013) has pointed out that the processes described in this framework are built on the L1 listening comprehension processes. Successful listening, which means the listeners have a clear concept of what the speaker intends to say, depends not only on linguistic processing (input decoding, lexical search and syntactic parsing), but also on higher-level processes (meaning and discourse construction). While lower-level processes enable the listeners to produce propositions and understand the literal meaning of the message being conveyed, higher-level processes assist the listeners in relating the incoming message to their existing knowledge and building a knowledge structure, resulting in a complete understanding of the message. To achieve complete comprehension, the listeners must engage in both levels of processing. Higher-level comprehension is not really possible if lower-order processes are not working efficiently.

2.2 Factors contributing to effective L2 listening

L2 listening comprehension appears to be restricted by two main factors: the level of listener's knowledge and the level of expertise or automaticity in processing (Buck, 2001; Field, 2013; Rost, 2002). The knowledge involved in language processing concerns both linguistic and non-linguistic knowledge. Linguistic knowledge or language-related knowledge is a domain of information in the individual's memory, and it is available for use in tandem with metacognitive strategies to create and interpret discourse in language use (Bachman & Palmer, 1996). Language knowledge comprises two broad categories: organizational and pragmatic. Organizational knowledge includes grammatical knowledge, knowledge of vocabulary, knowledge of syntax and knowledge of phonology/graphology (Bachman & Palmer, 1996; 2010). In listening processing, this type of knowledge is employed mainly in linguistic processing. It enables the listener to encode speech into linguistic units, detect phonetic features and recognize words in connected speech in order to

interpret the incoming text. Pragmatic knowledge is generally activated at a high level of processing, i.e. meaning and discourse construction (Field, 2013; Kintsch & van Dijk, 1978; Rost, 2002). It entails functional and sociolinguistic knowledge, both of which enable the listener to interpret text discourse by relating utterances or sentences to the speaker's intention and to the characteristics of the language-use setting (Bachman & Palmer, 1996). Another type of knowledge that affects L2 listening is the cultural or world knowledge that the listener brings to a listening situation (Field, 2013; Rost, 2002). Such knowledge is shaped by the listener's cultural background and experience. Similar to pragmatic knowledge, cultural knowledge is activated mainly in high-level processing (Field, 2013; Rost, 2002). The listener has to apply this type of knowledge, especially when he or she has to make inferences or references relevant to the message being delivered in order to understand its full and essential meaning.

Effective L2 listening depends not only on the listener's knowledge but also the degree to which he or she can process knowledge automatically (Field, 2013; Rost, 2002). As indicated in the previous section, listening ability integrates a number of psycholinguistic abilities working in a parallel and interactive manner. Rost (2002) divides these into four levels: neurological processing, linguistic processing, semantic processing and pragmatic processing, which Field (2013) categorizes into two levels of processing based on the level of cognitive development, i.e., lower-level and high-level processes. Lower-level processes entail linguistic processing consisting of decoding, word search and syntactic parsing while higher-level processes comprise meaning and discourse construction. Under normal circumstances, linguistic or lower-level processes are considered to be fundamental to listening and the skills that must be acquired prior to the development of higher processes, i.e. semantic and pragmatic or discourse processes (Field, 2013).

L2 processing occurs in association with automatic and controlled processing (Shiffrin & Schneider, 1977). Automatic processes are cognitive processes that are well developed and which put little or no demand on processing capacity (Shiffrin & Schneider, 1977). They do not require conscious attention and are therefore unavailable to conscious awareness (Shiffrin & Schneider, 1977). Controlled processes, on the other hand, are conscious. They require attention and are used flexibly in changing circumstances (Shiffrin & Schneider, 1977). In fact, what is necessary for complete text comprehension is automatic word recognition and syntactic analysis (Goh, 2002). When these lower level processes become automatic, more cognitive capacity is freed up for higher-level processing, such as making references and inferences and constructing meaning (Goh, 2002). If there is difficulty in processing a message at the level of linguistic processing, such as sound perception or word recognition, language users have little cognitive capacity remaining for higher-level processing, resulting in incomplete comprehension (Goh, 2002).

In conclusion, the literature indicates that there is a major problem in the use of integrated test tasks to assess L2 performance, which relates to a lack of clarity concerning the abilities assessed by the tasks. Particularly in the case of tasks that include listening input, i.e. listening-to-summarize tasks, it remains unclear what listening abilities are performed and measured. One way to investigate the construct or abilities underlying the language test task, as acknowledged by language testers (e.g., Bachman & Palmer, 1996; 2010; Messick, 1995) is to look into cognitive processes test takers used during the task performance. Cognitive processing, as pointed out by listening researchers (e.g., Buck, 2001; Field, 2013; Rost, 2002), depends upon several types of knowledge, including both linguistic (e.g., phonological, lexical, syntactic, semantic, pragmatic knowledge) and non-linguistic knowledge (e.g., topical and world knowledge). In addition, cognitive processing has been found to vary from one listener to another, depending on their competence and abilities to use

such knowledge to process for understanding. Investigation of cognitive processes and sources of knowledge used during integrated task performance, thus, can point to the construct or abilities assessed this task type.

The present study attempts to describe the listening construct underlying integrated listening tasks, e.g. listening-to-summarise tasks, by answering the following research questions.

a) What cognitive listening processes do ESL test-takers engage in while performing academic listening-to-summarize tasks?

b) Are there any differences in the listening processes involved when different language modalities, namely speaking and writing, are required for summary production?

3. Research methodology

3.1 Listening-to-summarize task materials

The task materials used in the present study comprise four test items adapted from PTE Academic.[1] Two tasks require participants to orally summarize a listening passage (listening-speaking tasks) and the other two require a written summary of the listening text (listening-writing tasks). The listening-speaking tasks were adapted from the *Re-tell Lecture* items, which originally ask test takers to retell what they have heard. The listening-writing tasks were taken from the *Summarize Spoken Text* items. An image related to the content of the listening input was added to each of the two tasks. The purpose of the modification is firstly to make the four tasks investigated comparable in terms of what test takers are supposed to do, and secondly, to study whether different language modalities (speaking and

---

[1] Pearson Test of English Academic – an English proficiency test for non-native English speakers who need to demonstrate their academic English ability for university admission or professional purposes.

writing) have an impact on listening processes. The listening input for each task is 60 to 80 seconds.

Strictly following PTE Academic guidelines, the participants are allowed to listen once only but they can take notes while listening. For the listening-speaking tasks, after listening the participants have 10 seconds to prepare and then 40 seconds to give their oral summary. For each of the listening-writing tasks, participants have 10 minutes to write a 50–70 word summary of what they hear.

3.2 Participants

Four Thai students at Lancaster University (one undergraduate and three postgraduates) participated in a pilot study. For reasons of anonymity, they are referred to in the finding section as P1, P2, P3 and P4. Based on their performance scores, participants were categorized into two groups: moderate scoring participants (P1, P3 and P4) and a low scoring participant (P2).

3.3 Data collection procedures

Data were collected on a one-to-one basis in the following order:

1) Completion of a background questionnaire

2) Completion of two sample listening-to-summarize items (one listening-speaking task and one listening-writing task), in order to familiarize participants with the item type and reduce test anxiety

3) Completion of four listening-to-summarize tasks, i.e. two tasks requiring an oral summary (listening-speaking) and two tasks requiring a written summary (listening-writing). These tasks were presented to the participants as a PowerPoint presentation (PPT), which was timed and set to play automatically when

participants clicked on the start button. A counterbalanced design was employed in the task delivery with the aim of minimizing the effects of task sequencing on performance. With this design, the participants began the test with different tasks.

4) Participation in a stimulated recall immediately after each task completion. This was carried out in the participants' first language, in this case Thai.[2] In each recall which took about 15-20 minutes to complete, the participants were first presented with the video recorded during their task performance. Then they were invited to explain what they were paying attention to or thinking about while listening.

3.4 Data analysis

The data were analysed as follows.

1) *Analysis of task performance*. Two experienced human raters scored all the task responses, using the human rater version of the PTE Academic scoring criteria, in order to evaluate performance level. The oral summaries were scored on three aspects: content, pronunciation and fluency. The written summaries were marked for content, grammar, vocabulary, form and spelling.

2) **Analysis of stimulated recall data**. The stimulated recall data were transcribed and analysed to identify cognitive listening processes. Following the notion of qualitative data analysis suggested by Gass and Mackey (2000), the data were categorized into episodes. The data were first segmented into what appear to be plausible units that correspond to Field's (2013) cognitive listening processes (see Section 2.1). For example, the following extract, obtained from one pilot study participant's protocol, was segmented into two chunks.

> [Chunk_1] When I heard 'handicraft', I told myself that it was about hand-made stuff, // [Chunk_2] but then it [the audio-recording] didn't say anything about

---

[2]The quotes in the findings section are translated from Thai.

items or products. Until I heard, 'his father', 'he', 'him', and 'the great scientist', I realized immediately that the listening was about a person [Hans Krebs], not about 'handicraft', as I had previously misunderstood.//

The first chunk (Chunk_1) indicates that the participant was trying to identify the word and its meaning she heard, which in this case she thought was 'handicraft'. This chunk was analysed to correspond to and categorized as lexical processing. The second chunk (Chunk_2) shows that the participants was trying to create a semantic relation between the words/phrases she heard in order to understand the real meaning of what she has been listening to. The participant linked 'his father', 'he', 'the great scientist' together and then realized that the audio was giving information about a person whose name was Hans Krebs, not about the hand-made as she understood in the beginning of the listening. This chunk corresponds to and was, thus, classified as semantic processing, i.e. reference making, in Field's framework.

4. Findings

Table 1 summarizes the cognitive listening processes demonstrated during the listening-to-summarize task performances obtained from stimulated recall data. Overall, the results show that the participants engaged in both lower- and higher- level processes to complete the tasks. Eight cognitive processes, in particular, were identified and categorized into three main types of processing: linguistic, semantic and discourse. Different language modalities (speaking and writing) required after the listening appeared to slightly affect the way the participants approached their listening tasks. Although the participants appear to employ the same processes, their performance scores show that they achieved different levels of success in their processing. Successful processing was found to depend, to a large extent, on the participant's linguistic and topical knowledge.

Table 1: Cognitive listening processes

Cognitive processes adopted by a proficient listener, as proposed by Field (2013)			Processes demonstrated by the participants			
Levels of processing	**Types of processing**	**Listening processes**	**Low scorer***	**Moderate scorers****		
			P2	**P1**	**P3**	**P4**
Lower-level processes	Linguistic	1) Input decoding	✓	✓	✓	✓
		2) Word search	✓	✓	✓	✓
		3) Syntactic parsing	✓	✓	✓	✓
Higher-level processes	Semantic	1) Identifying a speaker's purpose/context		✓	✓	✓
		2) Inferencing	✓	✓	✓	✓
		3) Referencing			✓	
	Discourse	1) Selecting				
		2) Integrating/linking pieces of information				
		3) Self-monitoring		✓		
		4) Structural building		✓		

*Average task performance score is lower than 40%

**Average task performance score is between 40% and 65%

1) Linguistic/lower-level processes

At the level of linguistic processing, three processes are found, namely input decoding, word search and syntactic parsing.

1.1) Input decoding

Input decoding, as described by Field (2013), is the lowest level of processing that takes place prior to word recognition. Words are recognized through the interaction of perceived sounds and context; and when listening to familiar words, this processing occurs automatically (Field, 2013). In this pilot study, only one participant (P1) explicitly indicated that he conducted input decoding. He describes, "here [the participant points to the video recorded while he was performing the task] I didn't know what the word was. I guessed from the sound I heard." The other three participants did not appear to engage in explicit input decoding. However, one can infer that they did use decoding, as they appeared to be able to

recognize words in connected speech as indicated in the next type of linguistic processing, i.e. word search.

1.2) Word search

All participants began their stimulated recall by describing that they were searching for key words while listening. For instance, P2 indicated, "I heard 'cells', 'human body', 'science' and 'study' and noted all the words down." P3 included in her notes, 'talent', 'really mean', 'high management', 'high ability', 'passenger happy' and 'I use the term to mean'. Then, from the words they recognized, they identified the points/ideas (propositions) in the listening task.

Although all the participants indicated that they started with word search when the listening began, it was found that moderate scoring participants (P1, P3 and P4) recognized words faster and more accurately than the low scorer, and as a result the moderate scoring participants were able to identify and infer the main points after listening to only a few sentences. P1, for example, said:

> When the listening started, I basically listened for vocabulary. Fortunately I recognized almost every word in the listening. I immediately understood it [the listening].

1.3) Syntactic parsing

The participants were found to adopt syntactic parsing for two purposes: to predict what was coming next in the listening and to build up propositions. Three participants (P1, P2, and P4) mention that they used syntactic parsing to predict what they were going to hear later.

P4 remarked:

> Here I was predicting that the speaker was going to talk about the definitions of talent because he said before, 'different ways of defining things restrictive, broad and meaningless'.

P2 reported:

> I know that he [the speaker] was going to talk about something in contrast because he said 'however… '.

In addition to syntactic parsing, all participants were found to make use of non-verbal information, (an image) provided as a supplement to the audio text, in order to recognize the words in connected speech and also to predict what the speaker was going to say next. For example, P1 commented:

> I predicted that the speaker was going to compare between the corruption and the income rate from the two graphs I saw while listening.

Two participants (P1 and P2) used syntactic parsing to build up their propositions. P1 explained:

> I have written in my note, 'as a result, ____ caused a disease'. I know that the missing word was a noun, so I put 'not having enough calcium in your blood' in the blank. It was a gerund phrase which I thought could function as a noun.

P2 used the word 'famous' in her proposition of 'he [Hans Krebs] is _____ for Krebs's cycle', because, as this participant said, "I know it needs an adjective and the adjective that often goes with for is 'famous'."

2) Semantic processes

Semantic processing, according to Field (2013), occurs when the listener is trying to understand the text beyond the literal meaning of the words uttered. It involves identifying the speaker's intention, inferencing and referencing, all of which were used by the participants.

2.1) Identifying speaker's intention

Field (2013) indicates that to comprehend a text, the listener also has to infer what the speaker leaves unsaid for whatever reason, e.g. believing it does not need to be included. The speaker's words are often insufficient to convey the full meaning of a message; and so the listener must identify what the real meaning of the message is. In this study, one participant (P1) was found to attempt to identify the speaker's intention in conveying a

message. P1 mentioned, "I think the speaker just wanted to add to the main point when he said 'no matter what parents say kids just do'."

2.2) Inferencing

On the basis of a few points that they could figure out, the participants started to identify the main points of the listening text. They did this by inferring from the words they noted down along with their background knowledge. For example, P2 stated:

> While listening, I had no idea what it [the audio text] was about at the beginning. Then, when I heard about 'cells', 'human body', 'science' and 'study', I assumed it was about a scientist.

However, it was found that the participant's topical knowledge caused one moderate scoring participant to misunderstand the story. As P4 stated:

> When I heard 'Hans Krebs' and 'he is a great scientist', I know that it was about a person. I studied about his life and work when I was in high school and I still remember his theory. I predicted that the story was about his life when he was a child and how he became famous.

It should be noted that, in the listening passage, Hans Krebs is only mentioned as an example of people who overcame obstacles and were successful in their life. This example is not the main point of the passage, but the participant misinterpreted it as the main point and thus scored "0" for content. This participant scored far less on this item than on the items she reported not having any content background on.

2.3) Referencing

Only one participant clearly demonstrated making use of referencing or linking reference words (e.g. he, him, this, what I just said, these factors) to their antecedents to construct the meaning of what was being said. This was stated by P3, one of the moderate scoring participants:

> When I heard 'handicraft', I told myself that it was about hand-made stuff, but then it [the audio] didn't say anything about items or products until I heard, 'his father', 'he', 'him' and 'the great scientist', I realized immediately that the audio text was about a person [Hans Krebs], not about 'handicraft', as I had previously misunderstood.

3) Discourse processes

Discourse processing occurs when the listener is constructing a discourse representation of what is being said. It is the highest level of listening comprehension processing (Field, 2013). Among the four types of discourse processing indicated by Field (2013), namely selecting, integrating, self-monitoring and structure building, the data obtained indicate only two types of processing, structure building and self-monitoring. Both of which were used only by P1, the moderate scoring participant who scored the highest in this study, in the two tasks that require an oral summary.

3.1) Structure building

The data indicated that P1 appeared to use this process only in the task that required an oral summary of listening. Because of the time constraints imposed by the oral summary task, this participant mentally outlined his summary while listening. He said, when watching his video, "here towards the end of the listening, I planned what I was going to say in the summary … from what I remember".

3.2) Self-monitoring

What is different between the highest-scoring participant (P1) and the lowest-scoring one (P2) is that, after they had predicted what the listening was going to be about, the lowest scorer (P2) listened and searched only for words she thought might help in constructing her predicted story. She was not aware that her predicted story could be wrong. The higher scorer (P1), though, was trying to predict and construct a mental outline, whilst also self-monitoring his own understanding by paying full attention to the rest of the listening, picking up on other key points and, realizing that his mental outline was not accurate, adjusting it. This participant (P1) said:

> At the beginning, I thought the speaker was going to describe the work of Hans Krebs so I planned to listen how Krebs's Cycle works, but then when the speaker

mentioned 'obstacles' and 'example of how people have overcome difficulty in life', I realized that the point he [the speaker] was making was about how people became successful in life rather than the work of a famous scientist. As you can see (this participant pointed to his video), I went back and corrected what I'd noted down before.

In sum, the stimulated recall data reveal that the participants activated several types of cognitive processing while doing the listening-to-summarize tasks. The processes identified correspond to three main types of processing in Field's framework, i.e. linguistic, semantic and discourse. At the level of linguistic processing, listeners decode, word search and syntactically parse. Semantic processing involves identifying the speaker's intention, inferencing and referencing. Discourse processing includes self-monitoring and structure building. Two types of discourse processing, i.e. selecting and integrating, were not shown in the stimulated recall data. Although the findings are presented in sequential order, it should be noted that, during actual processing, most of the time these processes occurred in parallel and were interactive.

5. Discussion

The results suggest that it is possible for listening-to-summarize tasks to tap into the higher levels of cognitive processing while listening and the processes necessary in academic listening contexts. That is, as they appeared to perform successfully on the tasks, the participants in this study had to engage in meaning construction and discourse processing. According to Field (2013), these are higher-level processes used by proficient listeners and necessary for success in academic studies. Although the tasks allowed the participants to employ higher-level cognitive listening processes, the low scoring participant appeared to process mainly at the linguistic processing level, perhaps due to linguistic knowledge limitations.

The different modalities (speaking and writing) required after the listening tasks appeared to affect the listening processes slightly. That is, the oral summary component appeared to force one participant to engage in structure building while listening, whereas writing summary did not. This is evident in P1's recall statement. P1 reported that he was aware of the lack of time to construct what he was going to say before having to speak, so he was structuring the content of his summary towards the end of his listening task. In this case, it might be possible that if the participant had more time to prepare after the listening as it was in the case of the writing summary, he might not have involved in structural building. However, because of the unique characteristics of an oral summary task that allow less than one minute to prepare before speaking, the participant was forced to engage in structural building while listening. The other processes tapped into by the tasks are however, in general, quite similar.

6. Conclusion

This study has sought to describe the construct or abilities underlying integrated test tasks or, more specifically, listening-to-summarize tasks. Knowing what exactly this task type assesses is crucial for interpretations and inferences made on the basis of test scores obtained from this item type. By conceptualizing the construct underlying the tasks as cognitive processes used by test takers, the study has investigated and revealed the processes that test-takers engage in during task performance. The analysis of stimulated recall data shows that listening-to-summarize tasks tap into three main types of processing, corresponding to Field's framework, i.e. linguistic, semantic and discourse processing. Moderate scoring participants, categorized according to their overall performance scores, were found to engage more in higher-level processes, i.e. meaning construction and discourse processing. These processes, according to Field (2013), are used by proficient listeners and are necessary for success in

academic listening. The low scoring participant, however, was found to process mainly at the linguistic processing level, perhaps due to linguistic knowledge limitations. In further investigation, it would thus be useful to find out what processes determine individuals' success in task performance and what sources of knowledge, e.g. linguistic or non-linguistic knowledge, individuals mainly rely on to complete tasks successfully.

The findings presented in this study rely exclusively upon stimulated recalls conducted with four ESL Thai participants studying towards their postgrad studies. There appear some concerns regarding the use of stimulated recalls. That is, in some cases listening processes occur automatically and participants may not be aware of the processes they have used. Another concern is that by the time the participants had completed the summary, they might not be able to clearly think back to the way in which they processed the listening text. Regarding these, other research methods are highly recommended to supplement stimulated recall data in future research. One of the methods that could be useful is an analysis of the summary content produced by test takers as it has been acknowledged to reveal processes activated in task completion (Johns & Mayes, 1990). In addition, since this study involved a small number of participants in one particular context and only one coder was used in data coding, the generalization of the results to different L1 background and learning contexts should, thus, be done with care.

References

Bachman, F. L., & Palmer, S. A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks (TOEFL Monograph Series MS-29)*. Princeton, NJ: Educational Testing Service.

Buck, G. (2001). *Assessing listening*: Cambridge University Press.

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: examining attributes of a free response listening test. *Language Testing, 15*(2), 119-157.

Chapelle, C. A. (1999). Validity in langauge assessment. *Annual Review of Applied Linguistics, 19*, 254-272.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing, 21*(2), 107-145.

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening* (pp. 77-151). Cambridge: Cambridge University Press.

Frost, K., Elder, C., & Wigglesworth, G. (2011). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test takers' oral performances. *Language Testing*, 1-25.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, N.J.: L. Erlbaum Associates.

Gebril, A. (2010). Bringing reading-to-write and writing-only assessment tasks together: A generalizability analysis. *Assessing Writing, 15*(2), 100-117.

Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly, 10*(1), 9-27.

Gebril, A., & Plakans, L. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks. *Assessing Writing, 21*(0), 56-73. doi: http://dx.doi.org/10.1016/j.asw.2014.03.002

Goh, C. C. M. (2002). Exploring listening comprehension tactics and their interaction patterns. *System, 30*, 185-206.

Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of University ESL students. *Applied Linguistics, 11*, 235-271.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

Lewkowicz, J. (1997). The integrated testing of a second language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education* (Vol. 7: Language testing and assessment, pp. 121-130). The Netherlands: Kluwer Academic Publishers.

Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes, 10*(2), 79-88.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing, 13*(2), 111-129.

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing, 17*(1), 18-34.

Rost, M. (2002). *Teaching and researching listening*. Great Britain: Pearson Education Limited.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*(2), 127-190.

Weigle, S. C., & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing, 21*(2), 118-133.

Weir, C. J. (1990). *Communicative language testing*. UK: Prentice Hall International Ltd.

Xi, X. (2008). Methods of test validation. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education:* (2 ed., Vol. 7, pp. 177-196): Springer Science and Business Media LLC.

Second language writing development from a Dynamic Systems Theory perspective

Attila M. Wind
Lancaster University, UK

Abstract

This study investigated how the lexical and syntactic features of two Hungarian advanced English as a Foreign Language (EFL) learners' writing evolved over a period of four months from a dynamic systems theory (DST) perspective. The participants provided data in the form of written essays which were analysed computationally by software packages (Coh-Metrix 2.0, Coh-Metrix 3.0 and Synlex L2 Complexity Analyzer). It was found that both lexical and syntactic indices showed interindividual and intraindividual variability. The log frequency for content words index showed a gradual decline which suggests that both participants started to use less frequent lexical items in their writing. When measure of textual lexical diversity (MTLD) was plotted against mean length of T-unit (MLTU) and MTLD against dependent clause per T-unit (DC/T), it was found that both participants concentrated on lexical complexity rather than on syntactic complexity which was also confirmed by the interview data. The largest rate change occurred for coordinate phrases per T-unit (CP/T) for both participants.

Keywords: *dynamic systems theory, second language writing writing development, syntactic and lexical complexity*

1. Introduction

Second language (L2) writing development has been investigated from various perspectives: dynamic systems theory (Verspoor & Smiskova, 2012), sociocultural theories of language learning (Wigglesworth & Storch, 2012), theories of multicompetence in language learning studies (Kobayashi & Rinnert, 2012), goal theories in education and psychology (Cumming, 2012), genre theories in second language writing research (Tardy, 2012), and systemic functional linguistics (Byrnes, 2012). However, Cumming (2010) points out that "no single theory might ever explain such complex phenomena as second language writing, which necessarily involves the full range of psychological, cultural, linguistic, political, and educational variables in which humans engage" (p. 19). Manchón (2012) argues that the same applies to the development of second language (L2) writing capacities since it is "intrinsically a multifaceted phenomenon that is mediated by a wide range of varied personal and situational variables" (p. 5).

Second language writing development, as with any development, is about change. Therefore, the obvious methodological design is longitudinal in nature (Ortega & Byrnes, 2008). Previous studies on second language writing have mainly employed cross-sectional designs (e.g., Bosher, 1998; Manchón, Roca de Larios & Murphy, 1998, 2000a, 2000b; Roca de Larios, 1996, 1999; Sasaki, 2000) which did not allow for the investigation of the changes in the development of individuals' second language writing. Cross-sectional studies used pretest-posttest designs which may not have been able to demonstrate development in writers' performance. In contrast, a longitudinal multi-wave research design (e.g., Berman, 1994; Sasaki, 2004), comparing written samples at more than two points in time, might be able to capture development.

Previous studies on L2 writing development have employed three different types of methods: (1) *quantitative* (e.g., Roca de Larios, Murphy & Manchón, 1999), (2) *qualitative* (e.g., Armengol-Castells, 2001; Cumming, Busch & Zhou, 2002), and (3) *mixed methods* designs (e.g., Cumming, 1989; Sasaki, 2002) to study L2 writing development. However, Norris and Manchón (2012) point out that the triangulation of data sources and analyses – using teacher and student interviews, classroom observations, and writing product analyses – may result in richer and more trustworthy interpretations. This study employed a mixed methods design triangulating the qualitative (semi-structured interviews) and quantitative findings (lexical and syntactic indices).

2. Theoretical underpinnings

In this section, the theoretical framework of this study is outlined: dynamic systems theory. In addition, recent views on lexical and syntactic complexity are reviewed.

2.1 Dynamic systems theory

Dynamic Systems Theory (DST), originally used to describe the behaviour of complex dynamical systems in applied mathematics, has been applied to several disciplines such as physics, biology and more recently to social sciences. In 1997 Larsen-Freeman published her oft-cited and pioneer work entitled *Chaos/Complexity Science and Second Language Acquisition* which made her the first researcher to study second language acquisition from a DST perspective. Larsen-Freeman (1997) characterised dynamic systems as "dynamic, complex, nonlinear, chaotic, unpredictable, sensitive to initial conditions, open, self-organizing, feedback sensitive, and adaptive" (p. 142).

One fundamental change in SLA terminology made by pro-DST applied linguists was to newly define the term *second language acquisition*. Long (1993) gives the following broad definition of an SLA theory:

> SLA theory encompasses the simultaneous and sequential acquisition and loss of second, third, fourth, etc. languages and dialects by children and adults learning naturalistically or with the aid of instruction, as individuals or in groups, in second or foreign language settings (p. 225).

De Bot and Larsen-Freeman (2011) broaden Long's definition even further by moving from *acquisition* to *development* and from *development* to *use*. They give the following definition: "a theory of SLD describes and ultimately explains the development and use of more than one language in individuals" (p. 6). The word *development* instead of *acquisition* in the definition refers to the fact that linguistic skills can grow and decline. In other words, language acquisition and attrition are equally possible outcomes of developmental processes. Moreover, *acquisition* suggests that at one point language is acquired, while *development* supports the belief that this process is ongoing.

Chaos theory studies the behaviour of dynamic systems which are highly sensitive to initial conditions. A small change at one point in a nonlinear system can result in large differences to a later stage. De Bot and Larsen-Freeman (2011) point out that in second language development "minimal differences between learners, even when they go through similar learning experiences, lead to very different learning outcomes" (p. 10). They simplify this statement by claiming that "similar teaching approaches do not necessarily lead to similar learning" (p. 10).

Every element is connected to other elements in a dynamic system. Obviously, a change in one system will have an effect on the other systems. For example, a change in the

learner's lexical system can affect the learner's phonological system. When learners acquire a new word (e.g. *born*) they have to learn its pronunciation (/bɔːn/) as well. Moreover, learners have to acquire its grammatical use (mainly in the passive) i.e. the acquisition of a new word also has an effect on the syntactical system.

Interaction with the environment and internal reorganisation result in a change in systems. An example of this feature in second language development is when language learners acquire a new word, usually a synonym (e.g. *gulp, sip* etc.) for a previously learnt more frequent word (e.g. *to drink*). In this case, the lexical system will reorganise itself by making differences between types of drinking (*sip, drink, gulp*).

In summary, DST is an ideal framework to study second language writing development. First, second language development is nonlinear in nature (Larsen-Freeman, 1997). Therefore, a two-wave (pretest-posttest) research design cannot plot individual growth trajectory. Instead, a multi-wave research design has to be adopted. Second, language systems (syntactic, lexical, phonologic, etc.) are interconnected. Thus, measuring only one of the language systems is not adequate to trace development. Instead, the investigation of more systems is necessary. Third, language development is dependent on initial conditions. Therefore, we need to collect as many different types of data as possible at the start of the study and during the investigation in order to discover how second language development took place.

2.2 Lexical complexity

Lexical complexity generally refers to lexical variability and lexical sophistication. The most reliable lexical variability measure to date is the measure of textual lexical diversity (MTLD) developed by McCarthy (2005) and validated by McCarthy and Jarvis (2010). MTLD is calculated as the mean length of sequential word strings in a text that are above a

certain threshold (0.72). MTLD calculates the type-token ratios (TTR) until the TTR falls to 0.72, when the first factor is produced. The counting of the TTRs is restarted. The final MTLD value is calculated by dividing the total number of words by the total number of factors. The calculation does not discard remaining data so a partial factor for remainders of the data is calculated. The programme runs forward and backward and the final MTLD value is obtained.

Lexical sophistication or lexical frequency profile might be indicative of a learner's vocabulary knowledge (e.g., Laufer & Nation, 1995). Lexical sophistication refers to the use of less frequent words in a text. Jarvis (2013) suggests assessing the overall commonness of the words in relation to the frequencies of those words found in large corpora such as the British National Corpus (BNC) or the American National Corpus (ANC). The most straightforward approach to measure lexical sophistication is to identify each word in the text with its rank in the corpus. The mean rank for all words would produce the lexical sophistication index. Another way to calculate lexical sophistication would be to convert rank orders to frequency bands (Laufer & Nation, 1995).

2.3 Syntactic complexity

Most studies on writing development have utilised quantitative measurements such as average length of structural units or the extent of clausal subordination. Researchers assumed that longer units and more subordination reflect greater complexity. A large percentage of these studies has relied on the construct of T-unit: "one main clause with all subordinate clauses attached to it" (Hunt, 1965, p. 20). The two most frequently used measures have been the mean length of T-unit (MLTU) (e.g., Larsen-Freeman, 1978, 1983; Ishikawa, 1995, Henry, 1996), which is the average across all T-units in a text, and clauses per T-unit (C/TU)

(e.g., Flahive & Snow, 1980; Bardovi-Harlig & Bofman, 1989; Hirano, 1991), which is the number of dependent clauses per T-unit.

The dependence on T-unit-based measures and clausal subordination was demonstrated in a review of literature by Wolfe-Quintero, Inagaki, and Kim (1998). Their extensive synthesis confirms that clauses per T-unit and dependent clause per independent clause have been the best complexity measures so far (pp. 118-119). The conclusion in the synthesis by Wolfe-Quintero et al. (1998) has made a huge impact on the studies of L2 writing development since its publication. Studies in the 2000s have relied heavily on T-unit-based measurements (e.g., Brown, Iwashita, & McNamara, 2005; Ellis & Yuan, 2004; Larsen-Freeman, 2006; Nelson & Van Meter, 2007) and subordinate clause ratios (e.g., Brown et al., 2005; Li, 2000; Norrby & Håkansson, 2007). In addition, Ortega (2003) confirms the heavy reliance on these two types of measurements in her survey of 27 studies. She found that 25 studies employed MLTU, while 11 studies relied on C/TU to measure grammatical complexity in college-level ESL and EFL writing. Other measures included were mean length of clause (MLC), mean length of sentence (MLS), T-unit per sentence (TU/S), dependent clause per clause (DC/C), dependent clause per T-unit (DC/TU), and clause per T-unit (C-TU).

The heavy reliance on T-unit-based and dependent clause measurements has received some criticism. For example, Bardovi-Harlig (1992) claims that "in evaluating the syntactic complexity of compositions written by advanced adult second language learners, T-unit analysis does not seem to reflect accurately the knowledge of the learner" (p. 391). The employment of these two measurements was also found problematic by several researchers in the 2000s (Rimmer, 2006, 2008; Ravid, 2005; Ravid & Berman, 2010; Norris & Ortega, 2009).

3. Methodology

This longitudinal case study investigated two Hungarian EFL learners' second language writing development – specifically the development of lexical and syntactic devices – by adopting the dynamic systems theory.

3.1 Research questions

This study answered the following two research questions.

1. Did the writers' performance evolve in terms of lexical complexity over a four-month period?

2. Did the writers' performance evolve in terms of syntactic complexity over a four-month period?

3.2 Research context

The present study was carried out in a private language school in Budapest, Hungary. This private institution offers language courses from A1 to C1 CEFR levels in four different languages (English, German, Italian and Spanish). At the school two different language exams approved by the Hungarian state can be taken. These are the language exam of the Budapest University of Technology and Economics and The European Language Certificates (TELC). The language institution offers courses which are specifically designed to prepare students for these two different language exams.

3.3 Participants

Two Hungarian EFL learners who were studying at the above-mentioned private language school participated in the study. The participants had to pass a successful language

exam at B2 CEFR level in order to be eligible to enroll for the course offered by the language school. The participants took part in 90-minute lessons twice a week from October 2012 to June 2013. The participants were the only students in the class and the course objectives were to develop students' four basic language skills i.e. writing, speaking, listening and reading skills. It is important to note that no one of the skills was emphasised more than the other during the course.

The first participant in the study, Augustine (a pseudonym), is a native speaker of Hungarian, aged 19. He took an English language exam at B2 CEFR level in 2012. Augustine started his university studies in 2012 at a university in Hungary. He has been learning English for more than 12 years. According to him, he has difficulty with grammar because he keeps forgetting the correct uses of tenses. He spent two weeks in London where he took part in a language course.

The second participant in the study, Andrew (a pseudonym), is also a native speaker of Hungarian, aged 18. He took a successful English language exam at B2 CEFR level in 2012. He has been studying English for more than ten years. He pointed out that his spoken English is worse than his writing. He spent one week in England with his family in 2009 and another week with his classmates in 2011. Table 1. shows a summary of the participants' profile.

Table 1.

Participants' Profile

		Augustine	Andrew
Gender		Male	Male
Age		19	18
L1 background		Hungarian	Hungarian
L2 learning	Length of learning English	12 years	10 years

experience	Length of staying in an English-speaking country	2 weeks	2 weeks
Level of English language proficiency		B2 CEFR	B2 CEFR

3.4 Instruments

Data for the research came from multiple sources and was collected by diverse methods. The primary data for the present study were participants' written argumentative compositions which were collected at 4 points (T1, T2, T3, and T4) over the 4-month period. I used four IELTS-type writing prompts (See Appendix 1.) which were taken from the IELTS Testbuilder (McCarter & Ash, 2003). The secondary data were the participants' responses during in-depth semi-structured interviews collected at two points (T2 and T4) over the 4-month period. The aim of the interviews was to find out more information about the participants' opinions, problems and difficulties in connection with second language learning, focusing especially on writing. I also wanted to find out how the participants form syntactic structures and how they choose a particular word to use. The secondary data served for triangulation purposes when commenting on the primary data.

3.5 Data collection procedure

The writing prompts were emailed directly to the participants so they could spend a week writing the essays. The request was to write four 250-word essays at four points in time responding to the given writing prompt. The participants could use dictionaries and spellchecker programs to help their writing. The dates of the interviews were chosen at the first meeting with the participants on 25th February in 2013. The two interviews were held at the school on 15th April and on 10th June. The interviews were semi-structured in order to allow flexibility and create a pleasant and relaxed atmosphere. The interviews lasted for 10-15

minutes in the participants' target language (English). However, I asked the participants to choose the language they felt more comfortable with in order to elicit as much information as possible.

Table 2.

The Time Frame of the Writing Prompts and Interviews

T1	March	Writing prompt 1	
T2	April	Writing prompt 2	Interview 1
T3	May	Writing prompt 3	
T4	June	Writing prompt 4	Interview 2

3.6 Measures of written performance

A comprehensive range of measures was selected to evaluate the syntactic features of students' writing (See Table 3.). The length of production unit was measured by the mean length of T-unit (MLTU). Sentence complexity was gauged by the Sentence complexity ratio (C/S). Subordination was measured by dependent clause per T-unit (DC/T). Coordination was measured by coordinate phrases per T-unit (CP/T). In addition, the complex nominals per T-unit (CN/T) index was also calculated. These measures were computed by Synlex L2 Syntactic Complexity Analyzer (Lu, 2010).

To measure lexical variability I adopted the measure of textual lexical diversity (MTLD), found to be the least affected by text length (Jarvis, 2012; McCarthy, 2005). MTLD was computed by Coh-Metrix 3.0 (Graesser, McNamara, & Kulikowich, 2011). Lexical sophistication was measured by log frequency of content words estimated by Coh-Metrix 2.0

(Graesser, McNamara, Louwerse & Cai, 2004) and based on the CELEX lexical database corpus.

Table 3.

Summary of the measures used in the study

Lexical Indices	Lexical variability	Measure of Textual Lexical Diversity (MTLD)
	Lexical sophistication	Log frequency of content words
Syntactic Indices	Length of production unit	Mean length of T-unit (MLTU)
	Sentence complexity	Sentence complexity ratio (C/S)
	Subordination	Dependent clauses per T-unit (DC/T)
	Coordination	Coordinate phrases per T-unit (CP/T)
	Particular structures	Complex nominals per T-unit (CN/T)

3.7 Data analysis

I had to deal with both macro- and micro-level perspectives. At the macro-level, quantitative measures were used to explore how the system changes and organises over time. At the micro-level, the participants' performance was examined from a qualitative standpoint.

Numerous software packages were used to analyse the syntactic and lexical features of the written argumentative compositions. These packages were Coh-Metrix 2.0 (Graesser et al., 2004) and Coh-Metrix 3.0 (Graesser et al., 2004; Graesser et al., 2011) and Synlex L2 Syntactic Complexity Analyzer (Lu, 2010).

4 Results

The different trajectories in Figure 1 clearly reflect the interindividual variability. Some individual performances show regression, progress, and others remain unchanged over time.

Individual growth in MTLD

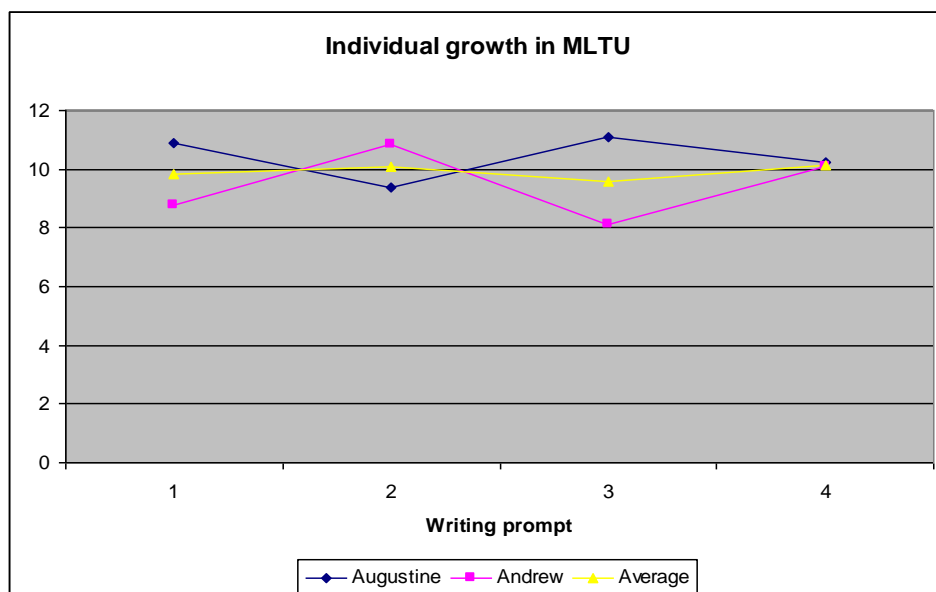**Individual growth in log frequency for content words**
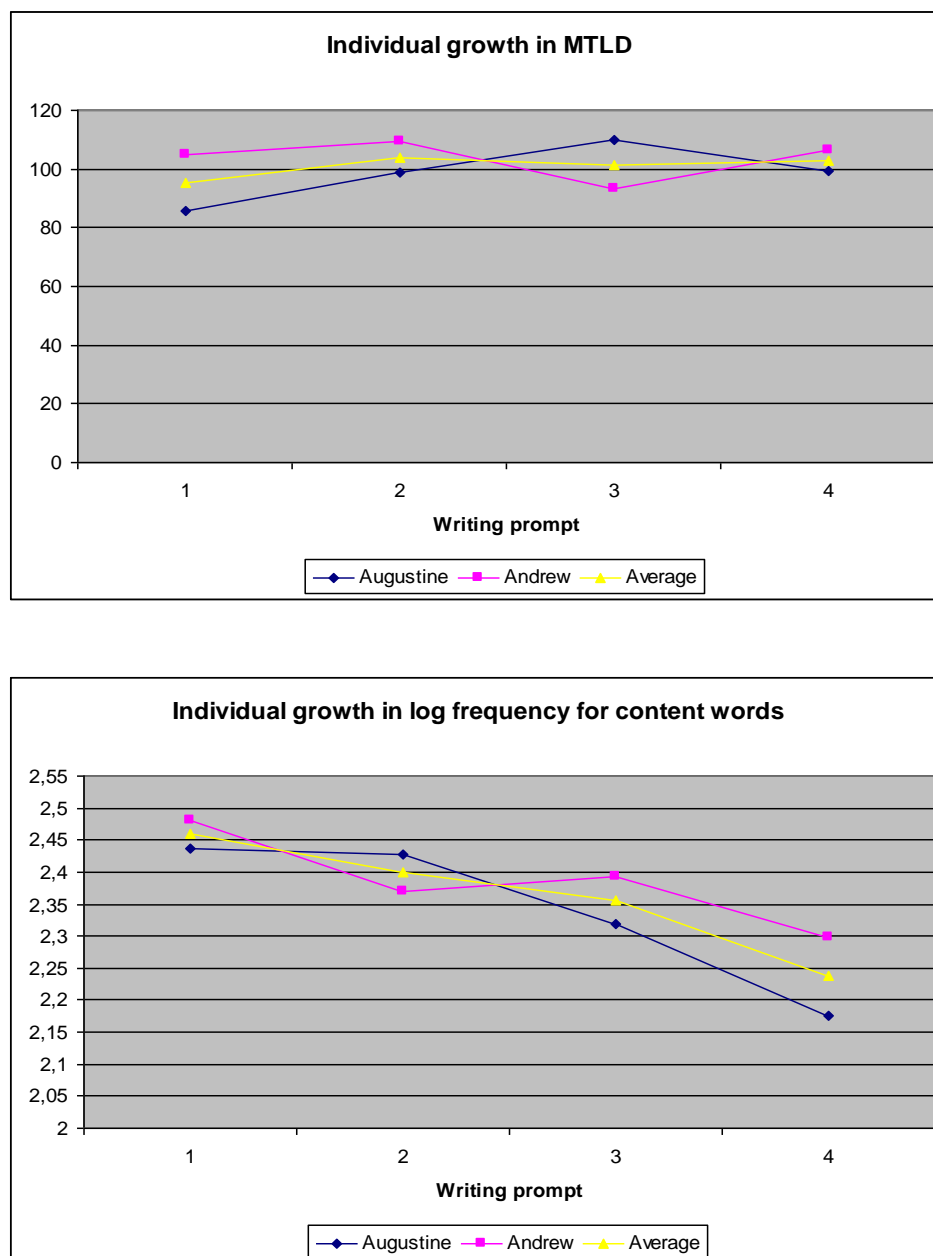
Figure 1. Interindividual variation over time on all indices
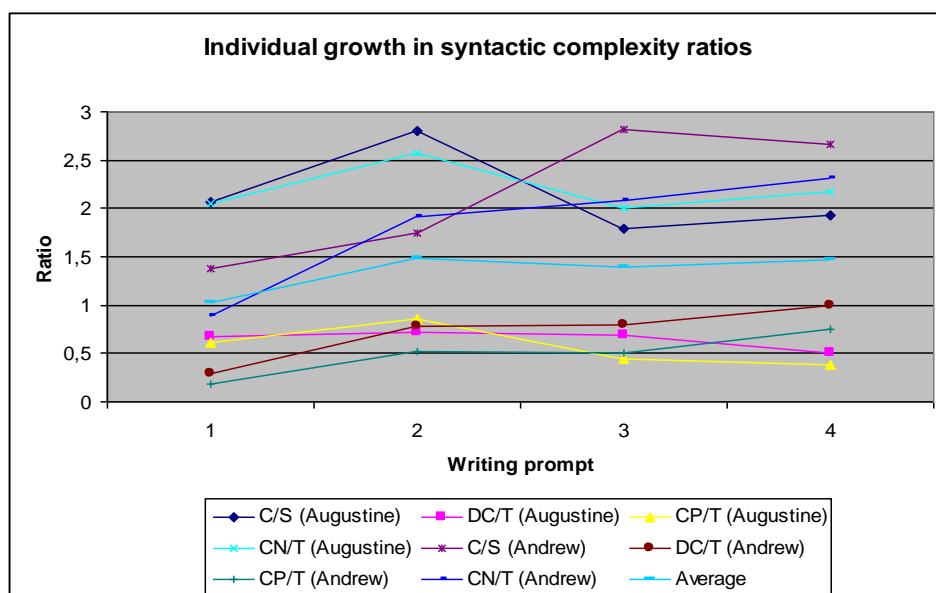
Figure 1. Continued

Whereas intraindividual variability was considered as a form of measurement error in traditional SLA analysis, from a dynamic systems approach, intraindividual variability is a fundamental source of information about the developmental process (van Geert & Steenbeek, 2005). The data were collected from another vantage point to highlight the intraindividual differences at differing data collection points. Figure 2. shows both participants' performance over time. The performance measures were transformed to z-scores in order to make the

comparability possible across the two indices of lexical complexity (MTLD, FREQ) and the five indices of syntactic complexity (MLTU, C/S, DC/T, CP/T, CN/T). Individual differences might be obscured by averaged data but averaged data within the individual might provide a true description of the behaviour of the individual (Sidman, 1960).
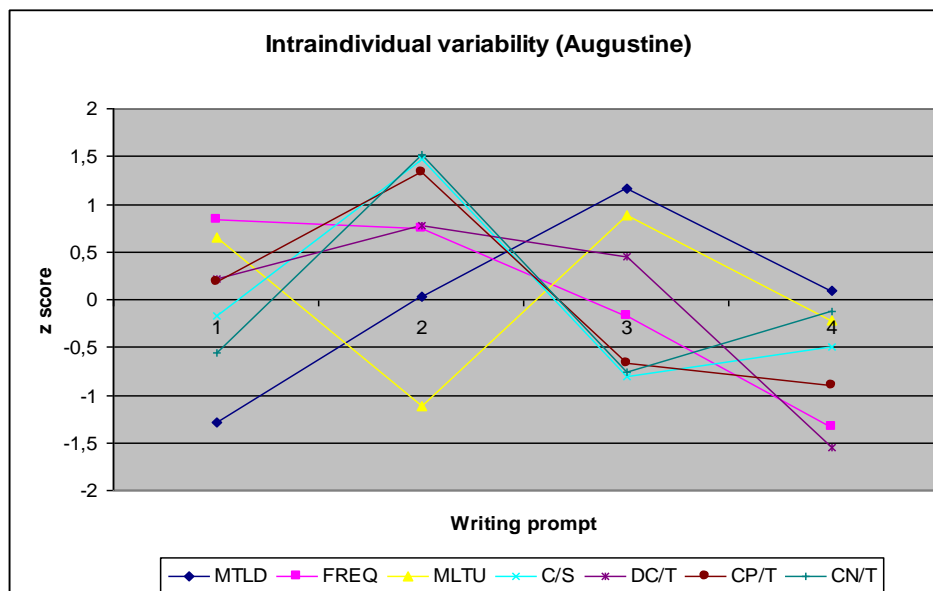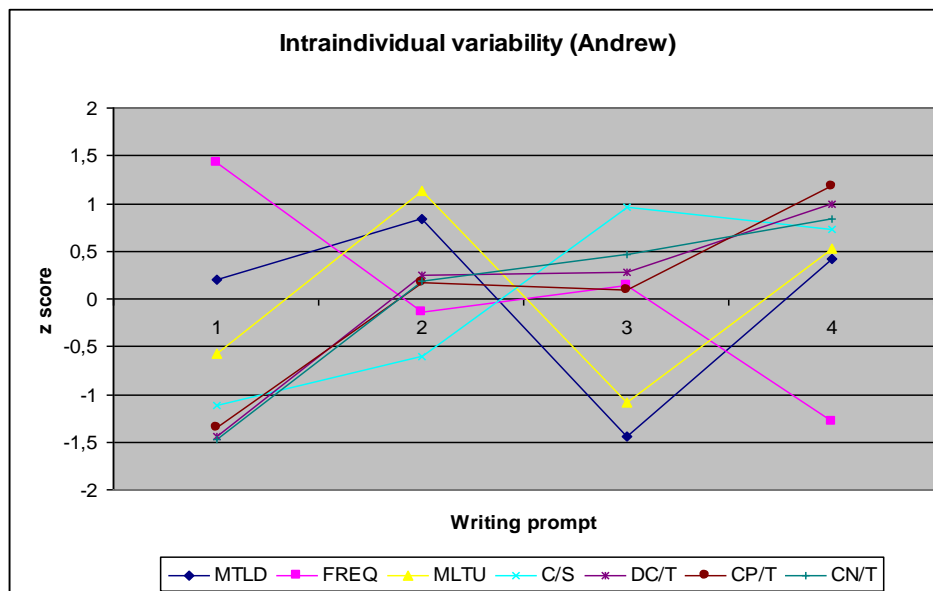
Figure 2. Intraindividual variation over time for both participants on all indices

While Figure 2. shows that there is intraindividual variability from one time to the next, Figure 3. demonstrates the identifications of attractors or preferred paths within

individual performances. The performance of both participants was mapped on two of the indices. When MTLD is plotted against MLTU and MTLD is plotted against DC/T (Figure 3.), it is clear that both participants have focused on lexical complexity rather than on syntactic complexity.
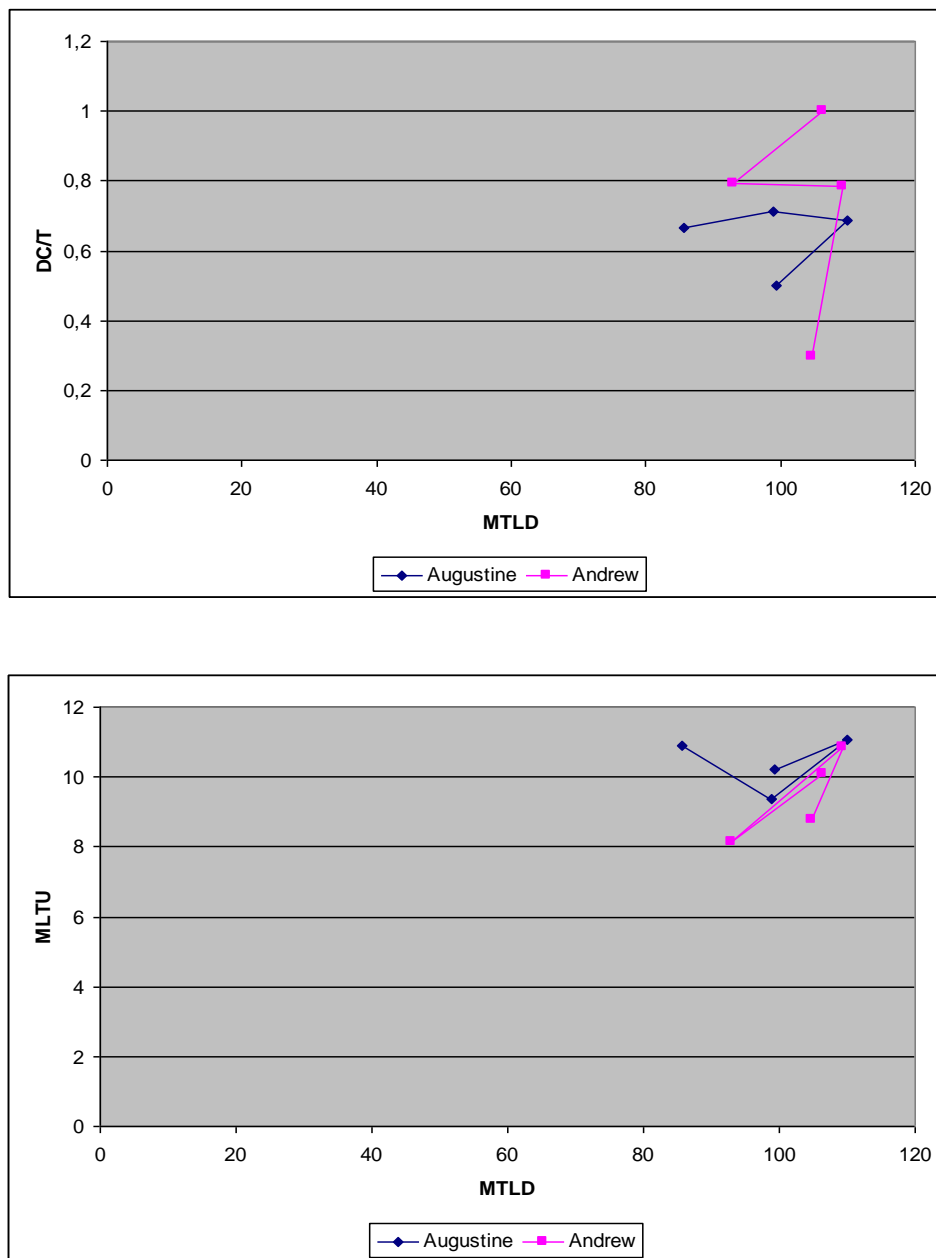
Figure 3. Change of MLTU compared with MTLD and the change of DC/T compared with MTLD for both participants

Since this is an exploratory study, techniques from L1 research were borrowed to display the same data from other perspectives. The rate of change over time was calculated

and this proved to be useful from a dynamic systems approach (e.g., Larsen-Freeman, 2006). The rate of change was calculated by setting the first data collection point (i.e. Writing prompt 1) as the baseline, with a value of 0. The rate of change is calculated by taking the difference between the next point of data collection (i.e. Writing prompt 2) and the previous point and dividing it by the previous data point. For example, as can be seen in Figure 1. Augustine's MTLD score at the first data collection point is 85.74. At the second data collection point, the score is 98.86, so the rate of change is $(98.86-85.74)/85.74 = 0.15$, which is plotted as the second data point in the graph, that is for Writing prompt 2. The results can be seen in Figure 4.

The rate of change fluctuates for both participants at different times. Figure 4. also reveals that the largest rate change occurs for CP/T for both participants. However, DC/T and CN/T also demonstrate a large rate of change in Andrew's graph at the second data collection point (Writing prompt 2).
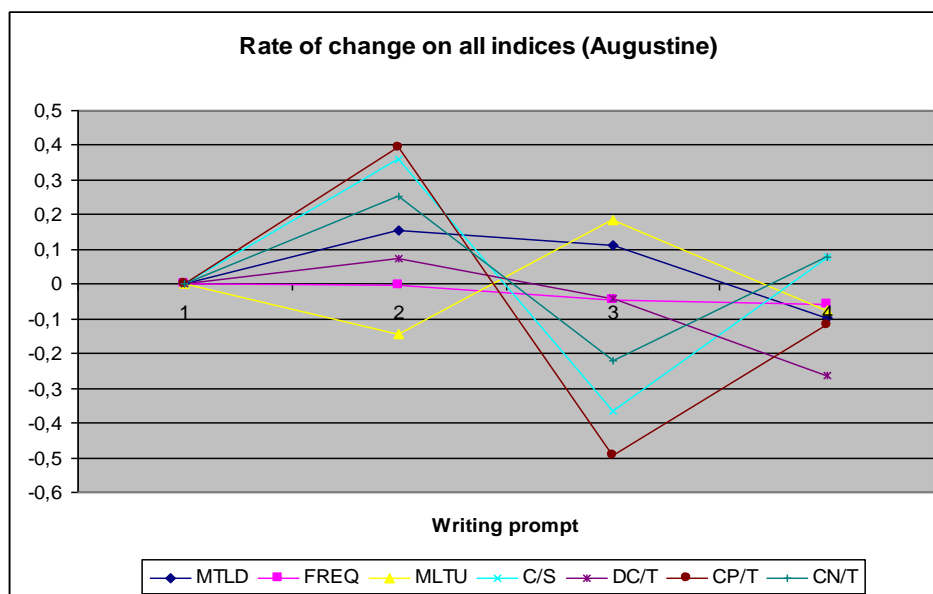

Figure 4. Rate of change on all indices for both participants over time
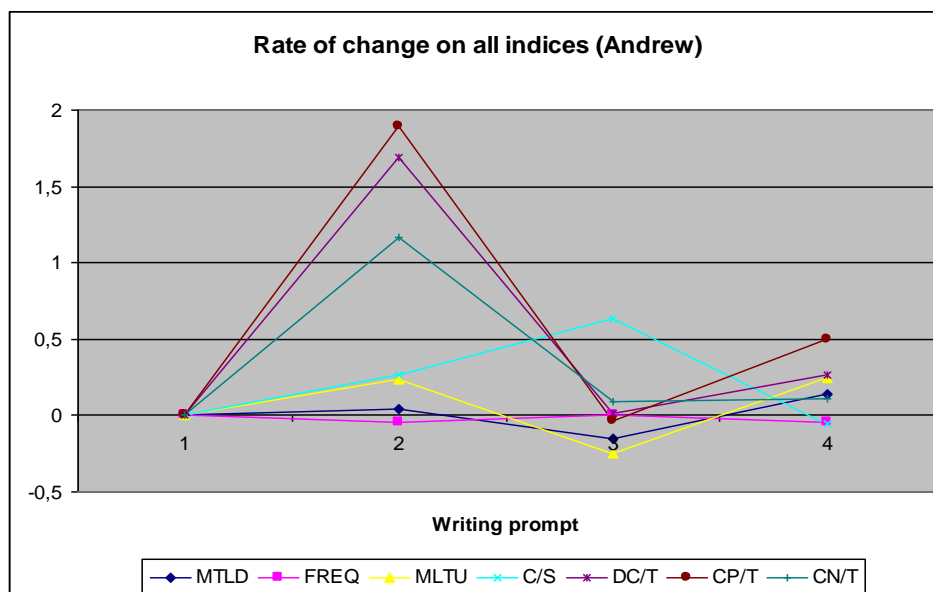
Rate of change on all indices (Andrew)

Figure 4. Continue

Both participants were interviewed twice at T2 and at T4. Before the interviews, a Lexical and Syntactic Complexity Profile (LSCP) was created which served as a stimulus during the interviews. Certain lexical items, which were deemed interesting, were highlighted, specifically less frequent words. Longer and more syntactically complex sentences were highlighted to explore the composing strategies the participants employed during the writing process.

The first set of data comes from Augustine, a 19-year-old Hungarian writer who has been learning English for 12 years. Table 4. contains lexical items and the original sentence structures taken from Augustine's written data. It was revealed that lexical items such as *exaggeration, emphasis, disregard, off the beaten track* were learnt from the coursebook he used during the course at the language school, while lexical items such as *ponder, cease, zealous* were taken from the online thesaurus. Table 4. also shows sentences taken from Augustine's texts. It was found that Augustine first thought about the sentences in his mother tongue and then he translated them into English. Therefore, his longer sentence structures seem unnatural to an English reader.

Table 4.

Augustine's Lexical and Syntactic Complexity Profile

Writing prompt	Lexical complexity	Syntactic complexity
1.	reckon exaggeration excessively ponder	Nowadays, a lot of people question whether computers are indispensable or hindrance. Moreover, youngers are really keen on sitting in front of their computers and chatting with friends but this aspect easily can be a problem if teenagers hanging on the computer for long hours, especially instead of learning.
2.	benefits or perks gratitude money emphasis crucial problem re-evaluate declination cease	It is a really common problem all around the world and in Hungary as well. In my view it is not a good solution that doctors get some bonus as a 'gratitude money', since other human beings get other benefits or perks, in addition they are usually provided fringe benefits, such as company company car, company computer or shopping coupons.
3.	deliberately intervene intimate proportion disregard deterrent zealous neutral	It is indispensable finding a balanced way of it. The main drawback is that many journalists are curious about the most intimate facts and moments of famous people and for the most part they do not respect human rights and write outrageous things and stories about stars, such as musicians, footballers.
4.	adolescent detrimental off the beaten track refrain	Third, the effects of drug abuse are well known. As a conclusion, the only reasonable way to solve this global issue to bring child up according to decent and acceptable habits, which would give a hand for them to acquire a good attitude, included not to prone to drugs, alcohol and other harmful habits.

The interview data collected from Augustine explained the stabilization in the MTLD and MLTU curves. Augustine understood the importance and the benefits of the writing

tasks. He was a motivated learner from the beginning. Therefore, he invested time and energy in the exercise. The excerpts from the interviews prove these findings:

Excerpt 1: 15 April (Augustine)

I found the writing exercise really useful since I wanted to take a C1 level language exam in the summer. I think by writing these essays I can really improve my writing skills. I used an online thesaurus to vary my vocabulary. I also wanted to learn new words and I thought using an online thesaurus would be really helpful.

Excerpt 2: 10 June (Augustine)

I mainly focused on grammar and vocabulary. The writing prompts weren't always easy. In some cases, I had to wait for minutes to able to write anything. It can be really stressful at exams.

The second set of data comes from Andrew, an 18-year-old Hungarian writer who has been learning English for 10 years. Table 5. shows Andrew's lexical and syntactic complexity profile. It was found that the lexical items *indispensable, garble, drastic measures, exaggerate* were acquired from the coursebook he used during the language course. Lexical items such as *aggregate, flagrant* and *albeit* were taken from an online monolingual dictionary, while the lexical item *uprising* was learnt from his teacher from his secondary school when they were told to talk about the Hungarian Uprising of 1956. Table 5. clearly demonstrates that Andrew improved his vocabulary over the four months because he started using less frequent lexical items in Writing prompt 3 and 4. This finding is confirmed by the quantitative measurements in Figure 4 (Log frequency for content words) where the declining curve shows that Andrew started to use less frequent lexical items. As far as

Andrew's syntactic complexity development is concerned, his longer sentence structures seem unnatural and in some cases (Writing prompt 2, 3, and 4) are incorrect in English. The interviews revealed that first he thought about the sentence structures in Hungarian and then he translated them into the target language.

Table 5.

Andrew's Lexical and Syntactic Complexity Profile

Writing prompt	Lexical complexity	Syntactic complexity
1.	investment indispensable engagement aggregate ignore	This investment simplified our lives. I think in the past like in the 18th or 19th century everyone was patient because they had to be that.
2.	garble undervalue extinguish reverence drastic measures uprising claim	Everyone likes music or movies or sport. Nowadays without any doubt, we can't live without doctors or nurses because for instance when we got ill or something disease no one knows better what medicines are essential or solution for our problems or when something very serious accident happens with ourselves there is no other way to be healthy again than a doctor save and expert treatment.
3.	disrespectful disturbance dismissive disregard notorious exaggerate sensationalist flagrant hindrance	They are must get used to being chased. Here it is a good example to confirm this statement: one of the members of the former band called Beatles mentioned that he has never known when he had the chance for to rub his ass because wherever he wanted to go at least a camera pursued him.
4.	albeit evolve compulsory anxiety abandon tremble endurance	Furthermore it takes the lead over their life. It means that the exaggerated amount of obligations that this generation get from their education or the compulsory work from their home might cross their endurance line and if they can't get rid of the feeling of being under pressure it provides them to try drugs.

The interview data explained the large rate of change (DC/T and CN/T) in Andrew's graph (Figure 4.) at the second data collection point (Writing prompt 2).

Excerpt 3: 15 April (Andrew)

When I had to submit the first essay, I felt it like a burden. I had many different tasks to do at school. Although I liked going to private English classes, I didn't want to do more exercises for extra classes, for example writing essays.

Excerpt 4: 10 June (Andrew)

When composing the third and the fourth essays I paid more attention to it. I took it more seriously. Augustine told me that these tasks could improve our writing. I realised that I might learn how to write essays in English which would be very useful at the language exam.

The qualitative data reveals that Andrew focused more on writing at the end of the study. He paid more attention to the composing processes.

To summarize, the DST approach revealed both interindividual and intraindividual variability over the four months. It was found that both lexical and syntactic indices showed variability. The log frequency for content words index showed a gradual decline which suggests that both participants started to use less frequent lexis in their writing. When MTLD was plotted against MLTU and MTLD against DC/T, it was found that both participants concentrated on lexical complexity rather than on syntactic complexity. The rate of change on all indices was larger in Andrew's text than in Augustine's.

5 Discussion

The goal of this study was to gain insight into the dynamic process of L2 writing development. The results are in line with the main features of the DST outlined in the previously.

I found stabilization when I averaged the progress of the two participants (Figure 1.) over the four-month-period. However, when I looked at the same data points for individuals, it was found that the lines were not stabilized at all. The lines sometimes went up and sometimes they went down. The graphs (Figure 1.) clearly show that the participants in this study followed different routes in SLA. Traditionally, such variability was considered as a form of measurement error in SLA. However, from a DST approach, variability is an essential source of information about the underlying developmental process. Although there is variability from one time to the next, attractors or preferred paths can also be identified within individual performances. When MTLD was plotted against DC/T, it appears that both participants worked on vocabulary at the expense of syntactic complexity (Figure 3.). This finding was confirmed by the interviews which revealed that participants focused more on vocabulary. In SLA, variability was explained by external sources. However, from a DST perspective a degree of variability cannot be explained by the effects of the external factors since some variability is an intrinsic and central characteristic of a self-organizing, dynamic system. The amount of variability constantly changes and that progress and regression follow each other, demonstrating nonlinear patterns of development.

The rate of change calculations shows that the system develops from an initial state and goes through iterations on the basis of available resources. As a consequence, there is no development without variability. The amount and type of variability can explain whether development took place or not. Therefore, it is essential to look at intraindividual variability

since it reveals the developmental dynamics which were ignored traditionally in SLA. The results show that even for an advanced learner, Augustine, the system can be far from stable. Although, a general increase over time is apparent for the index log frequency for content words, the development of all indices is nonlinear, showing moments of progress and regression.

6   Conclusion

This study investigated how two EFL Hungarian students' second language writing evolved over four months. It was found that both participants focused on vocabulary rather than on syntax. The DST perspective facilitated the interpretation of the interindividual and intraindividual variation over time. The multi-wave research design facilitated to investigate the changes in the syntactic and lexical systems of the participants. In addition, the mixed methods design facilitated the explanation of how the syntactic and lexical systems evolve over time.

In SLA it is quite frequent to compare macro-level group averages at different points in time. If reliable differences are found in mean levels of performance SLA researchers tend to conclude that development has occurred. The micro-level description of the individual's development is rarely addressed. In this study the graphs show that it is essential to ask if individual participants follow the same developmental pathways.

However, there are several limitations of this study. First, the four different writing prompts used in this study elicited different lexis which makes comparison difficult. Therefore, in future studies the writing prompts should be more carefully selected to control for the confounding effect of task characteristics. It is fundamental that the topic of the essays is controlled for because the topic might influence word choice and lexical measures. Second, in future studies the writing prompts should be written under controlled settings. The use of

dictionaries influences the lexical measures. If writers could not use dictionaries during the writing process, the actual vocabulary knowledge might be revealed. Third, it has to be noted that the syntactic complexity measurements used in this study - especially MTLU, DC/T – were found ineffective to measure the development of syntactic complexity in second language writing (Wolfe-Quintero, 1998; Ortega, 2003). Therefore, more genre-specific measurements are necessary (Biber, Gray, & Poonpon, 2011). For example, future studies should include ratios of conditional clauses, prepositional phrases, relative clauses, infinitive clauses, simple and complex postmodifiers which are characteristic of academic genre (Mazgutova & Kormos, in press). Fourth, lexical complexity cannot be measured by only two indices (MTLD and log frequency for content words) as in this study. Future studies should consider Jarvis's (2013) construct and include other dimensions of lexical diversity in the measurements such as evenness, volume, dispersion, or disparity. In this study only variability and rarity were measured. Finally, this study relied heavily on computational data analysis. Further studies should include human raters to measure syntactic and lexical complexity.

References

Armengol-Castells, L. (2001). Text-generating strategies of three multilingual writers: A protocol-based study. *Language Awareness, 10*(2-3), 91-106.

Bardovi-Harlig, K., & *Bofman*, T. (*1989*). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition, 11*, 17-34.

Berman, R. (1994). Learner's transfer of writing skills between languages. *TESL Canada Journal, 12(1),* 29-46.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*, 5-35. DOI:10.5054/tq.2011.244483.

Bosher, S. (1998). The composing processes of three Southeast Asian writers at the post-secondary level: An exploratory study. *Journal of Second Language Writing*, *7*, 205-241. DOI: 10.1016/S1060-3743(98)90013-3.

Brown, A., Iwashita, N., & McNamara, T. (2005). An examination of rater orientations and test-taker performance on English-for-academic-purposes speaking tasks. *TOEFL Monograph Series MS-29.* Princeton, NJ: Educational Testing Service.

Byrnes, H. (2012). Conceptualizing FL writing development in collegiate settings: A genre-based systemic functional linguistic approach. In R. M. Manchón (Eds.), *L2 Writing Development: Multiple Perspectives*. (pp. 191-220). Berlin, Germany: Walter de Gruyter.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33,* 497–505.

Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning, 39 (1),* 81-141.

Cumming, A. (2010). Theories, frameworks, and heuristics: Some reflections on inquiry and second Language writing. In T. Silva & P. K. Matsuda (Eds.), *Practicing Theory in Second Language Writing*, 48–71. West Lafayette, IN: Parlor Press.

Cumming, A. (2012). Goal theory and second-language writing development, two ways. In R. M. Manchón (Eds.), *L2 Writing Development: Multiple Perspectives*. (pp. 135-164). Berlin, Germany: Walter de Gruyter.

Cumming, A., Busch, M., & Zhou, A. (2002). Investigating learners' goals in the context of adult second-language writing. In S. Ransdell, & M. Barbier (Eds.), *New Directions for Research for L2 Writing*. Dordrecht, Netherlands: Kluwer.

Cumming, A., Rebuffot, J., & Ledwell, M. (1989). Reading and summarizing challenging texts in first and second languages. *Reading and Writing: An Interdisciplinary Journal, 2*(1), 201-219.

De Bot, K., & Larsen-Freeman, D. (2011). Researching Second Language Development from a Dynamic Systems Theory perspective. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.) *A Dynamic Approach to Second Language Development: Methods and techniques* (pp. 5-24). Amsterdam, Netherlands: John Benjamins.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26,* 59–84. doi:10.1017/S0272263104261034.

Flahive, D. E., & Snow, B. G. (1980). Measures of syntactic complexity in evaluating ESL compositions. In J. W. Oller, & K. Perkins (Eds.), *Research in language testing* (pp. 171–176). Rowley, MA: Newbury House.

Graesser, A. C., McNamara, D. D., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36,* 193–202.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. Educational Researcher, 40, 223-234. DOI: http://dx.doi.org/10.3102/0013189X11413260.

Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *The Modern Language Journal, 80,* 309-26.

Hirano, K. (1991). The effect of audience on the efficacy of objective measures of EFL proficiency in Japanese university students. *Annual Review of English Language Education in Japan, 2,* 21-30.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Champaign, IL: National Council of Teachers of English.

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4,* 51–70.

Jarvis, S. (2012). Lexical challenges in the intersection of applied linguistics and ANLP. In C. Boonthum-Denecke, P. M. McCarthy, & T. Lamkin (Eds.), *Cross-disciplinary advances in applied natural language processing: Issues and approaches* (pp. 50–72). Hershey, PA: IGI Global.

Jarvis, S. (2013). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Eds.), *Vocabulary Knowledge: Human ratings and automated measures* (pp. 13-45). Amsterdam, Netherlands: John Benjamins.

Kobayashi, H., & Rinnert, C. (2012). Understanding L2 writing development from a multicompetence perspective: Dynamic repertoires of knowledge and text construction. In R. M. Manchón (Eds.), *L2 Writing Development: Multiple Perspectives*. (pp. 101-134). Berlin, Germany: Walter de Gruyter.

Larsen-Freeman, D. (1978). An ESL Index of development. *TESOL Quarterly, 12,* 439-48.

Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H. W. Seliger & M. H. Long (Eds), *Classroom Oriented Research in Second Language Acquisition* (pp. 287-304). Rowley, MA: Newbury House.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, *18*(2), 141–165.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27,* 590–619. doi:10.1093/applin/aml029.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307–322.

Li, Y. (2000). Linguistic characteristics of ESL writing in task-based e-mail activities. *System*, *28,* 229–245. doi:10.1016/S0346-251X(00)00009-9.

Long, M. (1993). Assessment strategies for second language acquisition theories. Applied Linguistics, 14, 225-249.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*, 474-496. DOI:http://dx.doi.org/10.1075/ijcl.15.4.02lu

Manchón, R. M. (2012). Multiple perspectives in the study of L2 writing development: An introduction. In R. M. Manchòn (Ed.), *L2 Writing Development: Multiple perspectives* (pp. 1-16). Berlin, Germany: Walter de Gruyter.

Manchón, R. M. (2012). Multiple perspectives on L2 writing development. In, R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 17-46). Berlin, Germany: Walter de Gruyter.

Manchón, R. M., Roca de Larios, J., & Murphy, L. (1998). *Language ability, writing behaviors and the use of backward operations in L2 writing*. Paper given at the American Association for Applied Linguistics Conference, Seattle, WA.

Manchón, R. M., Roca, J., & Murphy. L. (2000a). An approximation to the study of backtracking in L2 writing. *Learning and Instruction, 10*(1), 13-35.

Manchón, R. M., Roca, J., & Murphy, L. (2000b). *The strategic value of backtracking in L2 writing.* Paper given at the AAAL 2000 Conference, Vancouver, Canada.

Mazgutova, D. & Kormos, J. (in press). Syntactic and lexical development in an intensive English for Academic Purposes programme.

McCarter, S., & Ash, J. (2003). *IELTS testbuilder*, Oxford: Macmillan.

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* [Microfiche]. Doctoral dissertation, University of Memphis.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment, *Behavior Research Methods, 42,* 381–392.

McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. *Institute for Intelligent Systems*, University of Memphis, Memphis, TN.

Nelson, N. W., & Van Meter, A. M. (2007). Measuring written language ability in narrative samples. *Reading and Writing Quarterly*, *23,* 287–309. doi:10.1080/10573560701277807.

Norrby, C., & Håkansson, G. (2007). The interaction of complexity and grammatical processability: The case of Swedish as a foreign language. *International Review of Applied Linguistics, 45,* 45–68. doi:10.1515/IRAL.2007.002.

Norris, J. M., & Manchón, R. (2012). Investigating L2 writing development from multiple perspectives: Issues in theory and research. In R. Manchón (Eds.), *L2 writing development: Multiple perspectives* (pp. 221-244). Berlin, Germany: Walter de Gruyter.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in SLA: The case of complexity. *Applied Linguistics*, *30,* 555–578. doi:10.1093/applin/amp044.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24,* 492–518. doi:10.1093/applin/24.4.492.

Ortega, L., & Byrnes, H. (Ed.) (2008). *The longitudinal Study of Advanced L2 Capacities*. New York: Routledge/Taylor & Francis.

Ravid, D. (2005). Emergence of linguistic complexity in written expository texts: Evidence from later language acquisition. In D. Ravid & H. Bat-Zeev Shyldkrot (Eds.), *Perspectives on language and language development* (pp. 337–355). Dordrecht, Netherlands: Kluwer.

Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language*, *30,* 3–26. doi:10.1177/0142723709350531.

Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing, 23,* 497–519. doi:10.1191/0265532206lt339oa.

Rimmer, W. (2008). Putting grammatical complexity in context. *Literacy*, *42,* 29–35. doi:10.1111/j.1467-9345.2008.00478.x.

Roca de Larios, J. (1996). Linearization strategies in L2 writing: Some observations. *Lenguaje y Textos, 8,* 191-208.

Roca de Larios, J. (1999). *Cognitive processes in L1 and L2 writing. A cross-sectional study.* Unpublished doctoral dissertation, University of Murcia, Spain.

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing, 9*(3), 259-291.

Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning, 54*(3), 525-582.

Sidman, M. (1960). *Tactics of Scientific Research*. New York: Basic Books, Inc.

Tardy, C. M. (2012). A rhetorical genre theory perspective on L2 writing development. In R. M. Manchón (Eds.), *L2 Writing Development: Multiple Perspectives*. (pp. 165-190). Berlin, Germany: Walter de Gruyter.

Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT Journal, 64*(3), 314-320.

Van Geert, P., & Steenbeek H. (2005). Explaining after by before: Basic aspects of a dynamic systems approach to the study of development. *Developmental Review, 25*(3–4), 408–42.

Verspoor, M. H., & Smiskova, H. (2012). Foreign Language Development from a Dynamic Usage-Based Perspective. In, R. M. Manchón (Eds.), *L2 writing development: Multiple perspectives* (pp. 17-46). Berlin, Germany: Walter de Gruyter.

Wigglesworth, G., & Storch, N. (2012). Feedback and writing development through collaboration: A socio-cultural approach. In R. M. Manchón (Eds.), *L2 Writing Development: Multiple Perspectives* (pp. 69-99). Berlin, Germany: Walter de Gruyter.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy & Complexity*. Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.

Appendix

IELTS-type writing prompts

Writing prompt 1.

You should spend about 40 minutes on this task. Present a written argument or case to an educated reader with no specialist knowledge of the following topic:

Some people believe that computers are more a hindrance than a help in today's world. Others feel they are such indispensable tools that they would not be able to live or work without them.

In what ways are computers a hindrance?

What is your opinion?

Use your own ideas, knowledge and experience and support your arguments with examples and with relevant evidence. You should write at least 250 words.

Writing prompt 2.

You should spend about 40 minutes on this task. Present a written argument or case to an educated reader with no specialist knowledge of the following topic:

Some people feel that certain workers like nurses, doctors and teachers are undervalued and should be paid more, especially when other people like film actors or company bosses are paid huge sums of money that are out of proportion to the importance of the work that they do.

How far do you agree?

What criteria should be used to decide how much people are paid?

Use your own ideas, knowledge and experience and support your arguments with examples and with relevant evidence. You should write at least 250 words.

Writing prompt 3.

You should spend about 40 minutes on this task. Write about the following topic:

Many newspapers and magazines feature stories about the private lives of famous people. We know what they eat, where they buy their clothes and who they love. We also often see pictures of them in private situations.

Is it appropriate for a magazine or newspaper to give this kind of private information about people?

Give reasons for your answer. Write at least 250 words.

Writing prompt 4.

You should spend about 40 minutes on this task. Present a written argument or case to an educated reader with no specialist knowledge of the following topic:

People in all modern societies use drugs, but today's youth are experimenting with both legal and illegal drugs, and at an increasingly early age. Some sociologists claim that parents and other members of society often set a bad example.

Discuss the causes and some effects of widespread drug use by young people in modern day society. Make any recommendations you feel are necessary to help fight youth drug abuse.

Use your own ideas, knowledge and experience and support your arguments with examples and with relevant evidence. You should write at least 250 words.

Authors

Paul Flanagan's work as presented here is based on a field trip to South Africa in summer 2013 as part of his doctoral research; he completed his PhD at Edge Hill University, Lancashire in May 2014. His research interests include syntax, language typology, corpus linguistics and lexical categories.

Bimali Indrarathne is a PhD student at the Department of Linguistics and English Language at Lancaster University. Her research interests are Second Language Acquisition, Task Based Language Teaching and English for Academic Purposes.

Anchana Rukthong is currently a full-time PhD student in Linguistics at Lancaster University. Her PhD research investigates cognitive processes and strategies used to understand listening input materials in listening-to-summarize tasks. For the past years, she worked as a lecturer at the Department of Languages and Linguistics, Prince of Songkla University, Thailand.

Attila M. Wind is a PhD student at Lancaster University. His research interests centre around second language acquisition (SLA); in particular, he focuses on second language writing development. He is interested in the role of motivational factors in writing. Attila is also an English and Italian teacher, language examiner, interpreter and translator. He has experience of teaching general English and Italian as well as EAP.

Editors

Bárbara-Pamela Olmos-López is a PhD student at Lancaster University. Her PhD research focuses on the analysis of authorial identity in academic discourse, particularly on undergraduate dissertations written in English as a Foreign Language. Her research interests are also on supervision, thesis writing, academic discourse, L2 writing and bilingualism.

Jing Huang is a PhD student in Linguistics at Lancaster University. Her research interests are Multilingualism, Critical Discourse Analysis, Language ideologies, Migration and Identity, Language policies, and Chinese Dialectology.

João Almeida is a PhD student at Lancaster University. His research project focuses on the relations between nationalist discourses and the sovereign debt normativity in Portugal's media during the 2011 bailout. He is currently researching about the relations between the economic and juridical dimensions of the 2007 crisis, drawing on Critical Discourse Analysis, Critical Theory, Political Economy of Communication, Constitutional Law and Political Philosophy.