Nordic Journal of Language Teaching and Learning

Development of metaphorical production in learner language: A longitudinal perspective

Susan Nacey

Inland Norway University of Applied Sciences

susan.nacey@inn.no

Abstract

This article details a longitudinal corpus-based exploration into the development of metaphorical production of L2 learners of English. The study tracks the progress of five secondary school pupils aged 13-17 in Norway, with the data consisting of texts written for end-of-semester exams: two texts per pupil over four consecutive academic years.

The overall goal is to shed light on how metaphorical production changes as pupils progress through different semesters and grades in their school careers. To do so, three subordinate aims are addressed. First, the study investigates how metaphor density varies over time, both for the group of pupils and for the individuals. In this regard, patterns for open-class versus closed-class metaphors across grade levels are also compared, to identify whether there is any particular level at which the use of the former overtakes the latter, as has been uncovered in previous research. A second aim is to examine the distribution of metaphor clusters over time, since clusters have been found to serve important discoursal functions and might therefore be expected to increase with improved proficiency over time. The third aim is to focus more closely on the identified metaphor clusters to explore the functions they serve in the written discourse of these language learners.

Keywords

Longitudinal study, learner language, Norway, metaphor, metaphor clusters, MIPVU.

Nordic Journal of Language Teaching and Learning, Vol. 10, No. 2, 2022. ISSN: 2703-8629
https://doi.org/10.46364/njltl.v10i2.975

1. Introduction

This article details an exploration into the development of metaphorical production of second/foreign (L2) learners of English, an area which is currently under-researched. The primary data for this study is longitudinal, thus adding a unique and innovative perspective to L2 metaphor studies. More specifically, the study tracks the progress of five secondary school pupils aged 13-17 in Norway, with the primary data consisting of texts produced for end-of-semester exams: two texts per pupil over four consecutive academic years. The empirical data has been retrieved from the the TRAWL (Tracking Written Learner Language) corpus, a collection of authentic L2 texts written by Norwegian pupils from grades five through the final year in upper secondary school (aged 10-18; see Dirdal et al., 2022). All metaphors in 40 texts were identified using the Metaphor Identification Procedure Vrjie Universiteit (MIPVU) (Steen et al., 2010).

The overall goal is to shed light on how metaphorical production may change as pupils progress through different semesters and grades in their school careers. To do so, three subordinate aims are addressed. First, the study investigates how metaphor density varies over time, both for the group of pupils as a whole, as well as for the individuals. In this regard, patterns for open-class versus closed-class metaphors across grade levels are also compared, to identify whether there is any particular level at which the use of the former overtakes the latter, as has been uncovered in previous research. A second aim is to examine the distribution of metaphor clusters over time, since clusters — that is, areas where metaphors appear in concentrated bursts within the texts — have been found to serve important discoursal functions and might therefore be expected to increase with improved proficiency over time. The third aim is to focus more closely on the identified metaphor clusters to explore the functions they serve in the written discourse of these learners. Note that in the present study, progression over semester and grade level, together with age, serves as a proxy for proficiency level. An underlying assumption is thus that L2 language proficiency grows with increased language instruction, exposure and learning opportunity.

Following this introduction, Section 2 presents the background for the study. It discusses why metaphor competence is important for L2 language learners, and presents an overview of previous research particularly relevant to the present study: Section 2.1 details research concerning the development of metaphor production as proficiency grows, while Section 2.2 discusses research investigating the function of metaphor clusters in discourse. Section 3 continues by presenting the primary material for the present study (Section 3.1), together with the methods used for metaphor identification (Section 3.2) and for the identification of metaphor clusters

(Section 3.3). Findings are then presented in Section 4, where each of the three subordinate aims are discussed in turn: metaphor density across semester, word class, and individual (Section 4.1), distribution of metaphor clusters across proficiency level per individual (Section 4.2), and function of metaphor clusters in written language of these learners (Section 4.3). Finally, the chapter finishes with concluding thoughts in Section 5.

2. Background

This investigation is grounded in Lakoff and Johnson's (1980) Conceptual Metaphor Theory, holding that metaphor — i.e., the phenomenon "in which we talk, and potentially think, about something in terms of something else" (Semino, 2008, p. 1) — is intrinsic to our conceptual systems and pervasive in language. Rather than merely being poetic elements, the metaphors we produce in language reflect the ways in which we (consciously or subconsciously) conceive of the world. The metaphors in our thought ('conceptual metaphors') are realized by metaphors in our language ('linguistic metaphors'). This process is so natural that most linguistic metaphors are conventional and codified in dictionaries, and language users need not be aware that they are 'doing metaphor'.

Being inherent to human nature, metaphor necessarily plays an important role in both L1 and L2 (or L3, L4, etc.) language learning. So-called 'metaphorical competence' — broadly defined as involving "the comprehension, awareness, and retention of metaphor in speaking, writing, reading and/or listening" (O'Reilly & Marsen, 2021, p. 26) — has been demonstrated to constitute an important component of all facets of learners' overall communicative competence (Littlemore & Low, 2006). Research about metaphor in L2 language and/or language learning has therefore focused on a number of different angles such as processing, comprehension, awareness, interpretation, etc. (see Hoang, 2014; Nacey, 2017 for overviews).

2.1 Development of metaphor production

The present study expands our knowledge base into L2 metaphor production, an area that started to come into focus only in recent years with empirical studies demonstrating the ubiquity of metaphor in L2 learner discourse (e.g. Nacey, 2013). As mentioned in the introduction, however, how metaphorical production *develops* as L2 proficiency increases is an especially under-researched area, with only a few published studies that have addressed this issue. Littlemore, Krennmayr, Turner, and Turner (2014) examined metaphorical use in 200 exam texts written by German and Greek learners of English: twenty texts for five proficiency levels per learner

group (levels A2 to C2 of the Common European Framework of Reference for Languages [CEFR]). They found an increase in metaphor density and in the number of metaphor clusters produced that corresponded with increased proficiency. Their work builds upon an unpublished MA thesis by one of the authors, who developed their 'moving metaphor density' method of calculating and visualizing metaphor density and identifying clusters in texts (Turner, 2010). His research shows that metaphor clusters are both more numerous, with increased metaphor density across proficiency levels. Hoang and Boers (2018), who mapped metaphor production in approximately 300 essays written by L1 Vietnamese learners of English from three different undergraduate years, similarly found a positive correlation between metaphor density and both year levels and grades obtained on the essays. Nacey (2019) reports similar results, finding that metaphor density and the number of instances of metaphor clusters increase as proficiency grows. This last study differs from the others in that the texts under investigation — which had also been collected in the TRAWL corpus, as in the present study — were produced by younger learners aged 5-19, rather than adults only. Cuberos, Rosado and Perera (2019) also include texts produced by younger learners, exploring metaphor production by L1 Arabic and L1 Chinese learners of Spanish across three age groups: grade school, junior high and university students. Although their primary focus is on deliberate metaphor, they too find that metaphorical competence in L2 discourse develops with age.

Littlemore et al. (2014) and Nacey (2019), in particular, both looked beyond the quantitative measure of metaphor density to explore the function of the observed metaphors as proficiency improves. Both found evidence that the role of metaphor evolves. At the lowest levels, metaphor serves little clear function and is mainly instantiated through function words such as prepositions and determiners. However, the role of metaphor grows more complex at later levels. Nacey (2019), for example, reports that by the upper secondary school level, metaphor contributes to a wide variety of textual features, including "(mostly) appropriate phraseology and collocations, more varied discourse markers, greater textual cohesion, personification, along with summing up and concluding" (p. 196). Note that all previous studies relied on *pseudo-longitudinal* data comprised of single observations (i.e. texts) produced by different learner groups who are in consecutive school grades or university levels or at different CEFR levels. By contrast, the present study examines *longitudinal* data comprised of repeated observations from the same informants over a time span of eight semesters.

2.2 Metaphor clusters and their functions

To conduct research into the function of metaphor at different proficiency levels, both Littlemore et al. (2014) and Nacey (2019) analyzed the use of *all* identified metaphors in the data, both metaphors that appeared within clusters and those that did not. Rather than looking into the general function of all metaphors in the longitudinal texts, by contrast, the qualitative exploration in the present study focuses primarily on the role of the identified metaphor clusters only. The reason for this focus is that despite observations of a positive correlation between proficiency and number of metaphor clusters, no one has yet explored whether such clusters are random or more purposeful in L2 learner language.

As a whole, however, there is general recognition that the distribution of metaphor in both spoken and written discourse is uneven (Cameron & Stelma, 2004, p. 108; Littlemore & Low, 2006, p. 135). The phenomenon of 'bursts' or 'clusters', i.e., stretches of text with a quantifiable increase in the number of metaphors, has been recognized since the 1970s (Pollio, 1977; Pollio & Barlow, 1975). Research into clusters has focused on the methodology of identifying metaphors and clusters, but also on the purpose of clusters — that is, where and why they occur. These investigations have explored L1 (English) discourse and indicate that the general functions of clusters are two-fold: either topical or organizational.

When it comes to topicality, Corts and Pollio (1999) observe that clusters tend to be relevant to the themes at hand, and are especially frequent when topics became "more complex, unfamiliar, or abstract" (p. 96). Corts (2006, p. 229) adds that such topical clusters tend towards both novelty and coherence, being derived from the same underlying metaphor. By contrast, Cameron and Stelma (2004), who observe many similar tendencies, also find many clusters that do not center around a single metaphor. They further observe that clusters are used for the explanation of difficult ideas and summarizing, while Cameron and Low (2004) also find clusters that have a mitigating effect or convey enthusiasm. Corts and Meyers (2002, p. 407) hypothesize that personal communicative style may at least partially account for the use of metaphor clusters, adding that they may be a mark of expertise, i.e. produced by people rich in knowledge and adept in explaining complicated concepts to laypersons.

When it comes to the discourse management function of metaphor clusters, many researchers point out that clusters often mark particular points of a discourse event; the start and end, topic changes, when intensive interactional work is happening in spoken discourse, and at paragraph changes in written discourse (e.g. Cameron & Low, 2004; Cameron & Stelma, 2004). Littlemore and Low (2006, p. 140) claim that such 'edge effects' have an intensification

purpose, as when a cluster towards the end of a text helps to generate a climax. One important consideration is whether definitive functions may be determined for all clusters: Corts (2006, p. 229) reminds us that an identified metaphor cluster "need not be coherent, novel or topical, rather it simply identifies a quantitative increase in production to a rate higher than normal."

3. Primary material and methods

This section details the material and methods employed in the present study. Section 3.1 first provides details about the texts analyzed for metaphor use. Section 3.2 then explains how metaphors were identified, while section 3.3 discusses the procedure for identifying metaphor clusters, which are subsequently explored to shed light on their potential role in the learners' texts. The overall methodology thus aligns with the suggestions in Deignan, Littlemore, and Semino (2013, p. 65): first identify figurative language, then calculate metaphor density, and finally determine whether the figurative language produced is associated with certain stages or transitions between stages.[1]

3.1 Primary material

The primary data for this study comes from the TRAWL corpus (see Dirdal et al., 2022). The overall ambition for TRAWL is the facilitation of studies into L2 and L3 language development across time, allowing for both pseudo-longitudinal investigations (i.e. a cross-sectional design where change in time is measured by a proxy, such as grade level or age; see Jarvis & Pavlenko, 2008, p. 37) and also for truly longitudinal studies where the production of individual learners is tracked through months and/or years. When it comes to English, TRAWL texts have been systematically collected from Norwegian classrooms in grades 5-13 (ages 10-19). These texts comprise naturalistic data, in that they have all been produced as part of regular coursework rather than at the prompting of researchers.

Corpus compilation began in 2015 and is ongoing, both in terms of text collection and subsequent preparation for inclusion in the final version of the corpus (e.g. digitalization, anonymization, coding of metadata, etc.). By the start of the present study, few sets of longitudinal data were available for research. Consequently, the primary material for the present study consists of texts produced by five pupils only, following their development in written English from

[1] All texts investigated here are available as supplemental material on https://dataverse.no/, along with all other documentation supporting the findings of this study: the data and R code used to generate the figures in this article, an overview of all exam prompts, general topics, identified metaphor clusters, and moving metaphor density charts for all texts.

grades 8 to 11. Grades 8-10 cover the entire three-year period of lower secondary school, when the pupils were in the same class. Grade 11 is their first year of upper secondary school; the five students had by then scattered to four different schools, with only two attending the same school together.[2]

The texts analyzed for the present study were all end-of-semester five-hour school exams, taken in October/November and April/May each year from the fall semester of 2015 to the spring semester of 2019. All the exams were modelled upon the tenth- and eleventh-grade national exams that are administered biannually. In the grades 8 and 9, the pupils were presented with adapted versions of national exams that had been given in previous years, ostensibly to gain experience with the exam format. Three of the texts produced in grades 10 and 11 were administered as mock exams based on exams from previous years, whereas the spring exam in the grade 11 was the actual national exam for that year. All five pupils sat for the same exams on the same days during all years.

The exams themselves, which are designed to test both topic knowledge and language ability, consist of two parts: a short-answer section focusing on the content of the study program and a long-answer section where the pupil is given a choice of (typically) four alternatives. The data for this study consists of the pupils' long-answer responses only. Prompts are intended to be interesting for the pupils, and frequently refer to preparation materials that are issued one day before the exam. The genre indicated in the prompts selected by these five pupils is either entirely unspecified (e.g. "write about" X) or underspecified (e.g. "write a text" about X), with the exception of a handful of times that pupils were instructed to write letters, 'informative' texts or 'creative' texts. As a result, the pupils were often free to choose the genre themselves and the texts are therefore varied, even when addressing the same prompt. They primarily include stories (adventure, science fiction, biographical), letters, and essays in grades 8-10. In grade 11, the pupils mainly produced fact-based articles and/or informative texts. Many texts fall into a 'school genre', "where the student is invited to display knowledge about syllabus-related topics while drawing on his/her personal experience or opinion" (Ørevik, 2012, p. 8). An example is seen in the following prompt from the grade 10 fall exam, which two of the five pupils addressed:

[2] Rørvik's study (2022), which investigates the longitudinal development of noun-phrase complexity, also includes these five pupils as informants. For the purposes of comparison, please note the following correspondences between the pupils in the present study and those in Rørvik's: pupil A = P102, pupil B = P103, Pupil C = P105, Pupil D = P109, Pupil E = P110. Note also that in the TRAWL corpus, all pupils were assigned a longer identification tag: this study's pupils A-E correspond to pupils P60102, P60103, P60105, P60109 and P60110.

> Of the traditions and lifestyles described in the reference material or that you have studied in your English class, explain which country you would live in if you could choose. Give specific reasons as to what traditions in the country you would like to experience and why.

Topics have varied over the years, including hobbies, geographical locations, historical figures, personal relationships, the importance of English in the world, and the influence of celebrities.

A total of 26,780 lexical units were analyzed for metaphorical use.[3] The number of lexical units varies per pupil and grade level, with the longest texts having been produced in grades 8 and 9 as various types of stories. The boxplots in Figure 1 provide a more detailed overview.

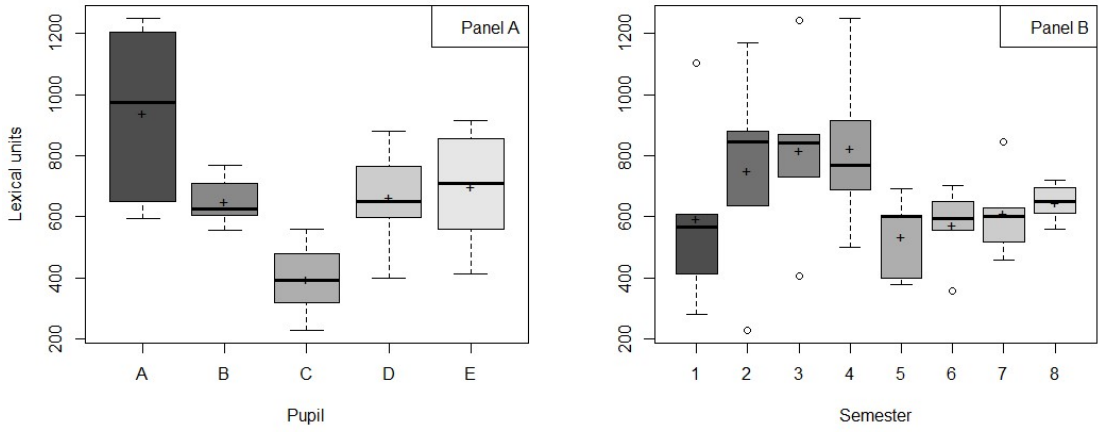

Figure 1: Text length per student (Panel A) and semester (Panel B).

Panel A shows the distribution of text length for each of the five pupils (A-E). Here we see that Pupil A was the most prolific, having produced 7505 words in total; Pupil C wrote the least, with 3171 words (median = 5309, mean = 5356). Panel B, by contrast, shows the distribution of text length over time. Each set of exam responses corresponds to one semester, ranging from semester 1 (grade 8 fall semester) to semester 8 (grade 11 spring semester). Semesters 1 and 2 thus correspond to grade 8, semesters 3 and 4 to grade 9, semesters 5 and 6 to grade 10, and semesters 7 and 8 to grade 11. Text length per semester over these four years ranges from a minimum of 2675 words in semester 5 to a maximum of 4123 words in semester 4 (median =

[3] Note that the unit of analysis for the metaphor identification procedure employed in the present study is the lexical unit. This corresponds to the orthographic words in all cases except for polywords (e.g. *of course*), compounds (e.g. *pop star*) and phrasal verbs (e.g. *wake up*). Following established practice, the terms 'lexical unit' and 'word' are henceforth used synonymously in this article (see Steen et al., 2010, pp. 26-32).

3147, mean = 3348). Overall, the texts range in length from 229 words in Text 2C to 1250 words in Text 4A (median = 634, mean = 669.5).

3.2 Metaphor identification

The Metaphor Identification Procedure Vrije Amsterdam (MIPVU; Steen et al., 2010) was employed to identify all linguistic metaphors in the 40 texts; underlying conceptual metaphors were not identified. MIPVU identifies 'indirect' and 'direct' metaphors.[4] Indirect metaphors consist of those words where there is a contrast between the contextual sense and a more basic sense, and where that contrast may be attributed to some form of comparison. As an example, consider the italicized word *consume* from a grade 10 text about the advantages and disadvantages of online and offline connections.

(1) ... phones and computers *consume* more and more of our time ... (6D)[5]

The most basic sense of *consume* (that is, the most concrete, specific, and human-oriented sense in the dictionary) is "to eat or drink something", the first sense entry for the verb in the online version of the *Macmillan Dictionary*.[6] By contrast, the verb's contextual sense is its fourth sense entry, "to take all of your attention so that you cannot think of something else". These two senses are sufficiently distinct (i.e. represented by different sense entries) and may be understood by comparison: we understand the undivided interest in something in terms of eating or drinking.

In direct metaphors, an underlying comparison is expressed through 'direct' language use, where there is no contrast between the contextual and more basic sense. An example is the simile in 0 from a science fiction story in a grade 8 text.

(2) ... she slept *like a rock*. (1A)

Here there is no distinction between the contextual and basic senses of *rock*, even though there is an underlying conceptual metaphor because the discourse topic here is the manner of

[4] MIPVU also identifies 'implicit' metaphors where the metaphorical relationship is due to an underlying cohesive grammatical link referring to metaphorical text, as is the case with e.g. anaphoric pronouns or determiners. Although they were identified in the current study and thus contribute to calculations of metaphor density, they are of lesser importance for the present study and not discussed.

[5] All quotations from the pupils' texts are followed by a code identifying the text where it is found (numeral = semester 1-8, letter = pupil A-E). These quotations are reproduced as written, including any mistakes/errors.

[6] https://www.macmillandictionary.com/

sleeping, rather than rocks. The evocation of an 'alien' physical source domain that is unrelated to the actual topic thus requires the reader to establish a cross-domain comparison between the referents of the words in discourse. Direct metaphors are sometimes signaled by 'metaphorical flags' such as *like* in 0, but they may also appear without any such flags, as often happens with metaphorical analogies.

3.3 Identification of metaphor clusters

This study adheres to a clear operational definition of metaphor clusters, identifying them through a series of time analyses, one per text, following Turner (2010) and Littlemore et al. (2014). This procedure calculates the metaphor density of overlapping 20-word spans, starting with a calculation of the average metaphor density for the first 20 words (the number of identified metaphors divided by 20), then shifting one word for the next 20-span calculation, and continuing in this manner until the end of the text is reached. The results of each analysis may then be graphically illustrated by a figure showing the moving metaphor density of the text. An example from Text 6A is provided in Figure 2.

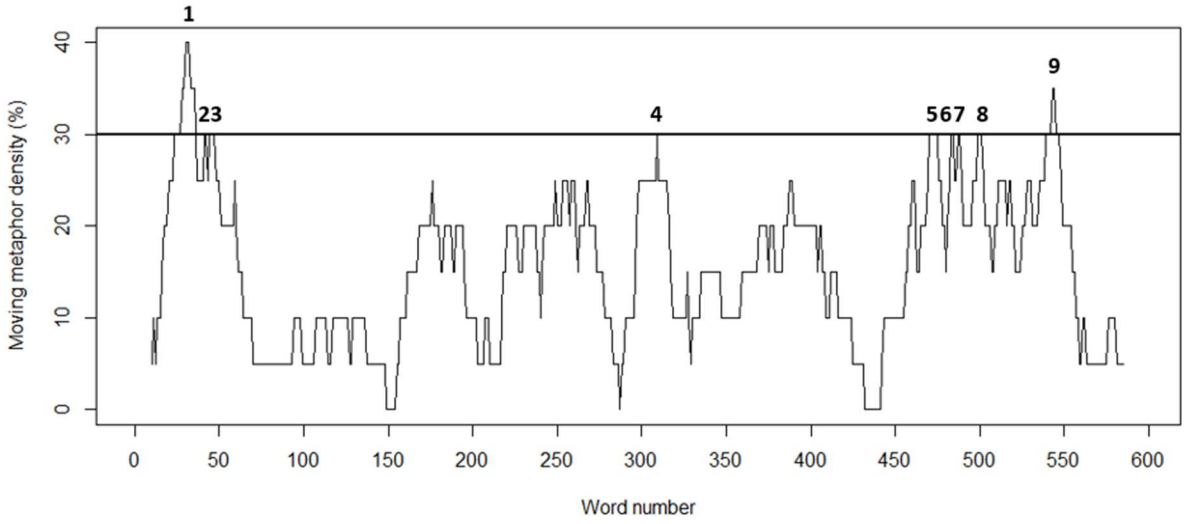

Figure 2: Moving metaphor density chart: Text 6A.

The cut-off point for what is considered a metaphor cluster is 30% (indicated by the black horizontal line in Figure 2), which Littlemore et al. found to include "visible metaphor above and beyond the sorts of highly conventional metaphorical uses of prepositions" (2014, p. 124). In this way, nine clusters were identified in the text represented by Figure 2. These are numbered above each peak.

As an illustration, extract 0 shows the text corresponding to peak 8. Following Turner (2010), the span length is indicated by forward slashes and the identified metaphors in the cluster are italicized.

> (3) Now, here is the / question: Is *having access to* the Internet the *key to* **learning English? Well,** if it is the *key* is quite *hard* to / say.

Here we see that the total length of the entire cluster exceeds 20 words because three consecutive 20-word spans exceed 30% metaphor density: the respective mid-points of each span are highlighted in bold in the extract. Figure 2 shows that some such consecutive spans appear as even wider 'plateaus' (e.g. peak 1), while some clusters are short but have overlapping spans near each other (e.g. peaks 2 and 3). The moving metaphor density charts essentially provide individual metaphor portraits for each text through clearly illustrating individual textual variation in metaphor use far beyond what a simple calculation of average metaphor density for an entire text shows.

4. Findings

This section discusses the findings related to the each of the study's three subordinate research aims. Section 4.1 begins by detailing the metaphor density of the texts per semester and word class, both for the texts as a whole as well as for the individuals. This section also provides text extracts with varying degrees of metaphor density, to demonstrate how increased density may be realized in practice. Section 4.2 continues by focusing on the observed occurrences of metaphor clusters on both the group and individual levels. Finally, Section 4.3 looks at the function of metaphor clusters across the semesters. Two subsections explore the functions of the observed metaphor clusters in grades 8 and 9 (Section 4.3.1) and in grades 10 and 11 (Section 4.3.2).

4.1 Metaphor density per semester, word class, and individual

In general, metaphor density varies widely in the texts, ranging from a minimum of 2.9% in Text 1E to a maximum of 19.3% in Text 8D (median = 10.4%, mean = 10.5%). The scatterplot in Figure 3 presents the mean metaphor density per text and semester. Each dot represents one text (40 in total), with a 'jitter' function having been applied to separate otherwise overlapping dots. The vertical axis shows the metaphor density of each text, while the horizontal axis indicates the semesters in chronological order. Semesters 1 and 2 thus correspond to the grade 8

fall and spring semesters (respectively), and the sequence continues through semesters 7 and 8, representing the fall and spring semesters in grade 11.

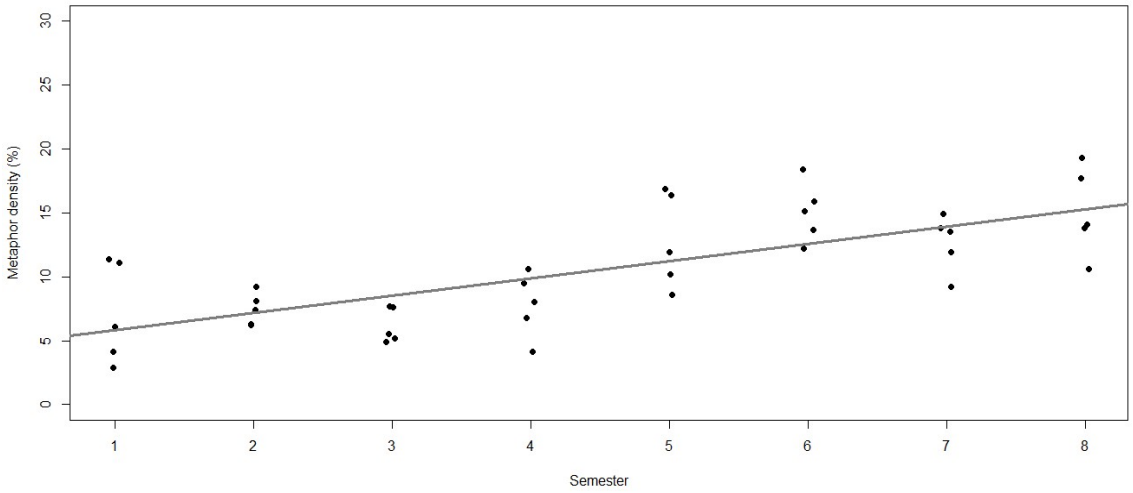

Figure 3: Metaphor density per semester.

This figure indicates a positive association between metaphor density and semester (V=15, $p<0.05$). As the independent variable of semester increases, so too does the dependent variable of metaphor density. The slope of the trend line indicates that we can generally expect metaphor density to increase by an average of 1.3% per semester (95% CI = 0.9, 1.8). This finding aligns with those in previous research, providing further evidence that increased metaphor density may be one characteristic of increased language proficiency.

Extracts 0 through 0 illustrate how increasing metaphor density may be realized in practice. These extracts are all retrieved from Text 5B, the text reaching the maximum metaphor density in the data for that pupil. For this essay, written for the grade 10 spring exam, the pupil had been asked to reflect upon the relationship between Dina and Nusswan, the two protagonists of a short story in the exam's preparation material, with regard to Dina getting married (a marriage forced by her brother Nusswan). These extracts show four 20-word spans that rise in metaphor density by 10% increments, from 0% metaphor density in 0 to 30% metaphor density in 0. This last extract, which reaches the 30% threshold inherent in the identification of metaphor clusters, is thus part of the first of this text's two clusters (illustrated in the moving metaphor density chart presented further on in Figure 7, section 4.3.2). As with extract 0, each 20-word span is demarcated by forward slashes, the mid-points are in bold script, and identified metaphors are italicized.

(4) / He asks if it`s Solly or Porus. She says no. **Then** he asks if it`s Dara or Firdosh. She / says no. (5B; 0% metaphor density)

(5) After several years of helping Nusswans / wife *with* the household chores, she is now ready **to** do them *on* her own in her own house. / I think she... (5B; 10% metaphor density)

(6) Nowadays it is difficult to / *have* your own opinions because you get to *see* what **everyone** else think. The *pressure* to *have* a "perfect" body / is a very common problem... (5B; 20% metaphor density)

(7) but the / parents should be *giving tips about* the person their child **is** dating. You could almost *make* a *parallel to* slave / buying and selling. (5B; 30% metaphor density)

Extract 0, containing no metaphor, recounts a particular scene from the short story, the type of concrete happening that does not naturally trigger metaphor use. As noted in section 3.1, such depictions are most common in texts from the earlier grades, many of which are stories. The growth to 10% metaphor density that we see in 0 shows an incremental increase that may be due to the metaphorical role of function words. Here the two metaphors in the 20-word span are both prepositions, which is the most metaphorical word class (Nacey, 2013, p. 155). In extract 0, where the metaphor density has reached 20%, we find both lexical and function words. Finally, extract 0, showing a span with 30% metaphor density, illustrates how metaphor may contribute to more complex phraseology and collocations, such as *give tips about X.* Although not all may necessarily conform to typical patterns in L1 English (e.g. *make a parallel to X* would more commonly be expressed as *draw a parallel between*), metaphorical phrases grow in length and number.

Although metaphor density thus increases in the texts as a whole, we see variability in outcomes over time when we look at the individual level. While the metaphor density of all five pupils increases over time, it does so at different rates. These rates vary from 0.5% for Pupil C to 2.1% for Pupil A, meaning that the maximum rate of increase is roughly four times the minimum rate. While such difference is in and of itself unsurprising since we know that individual behavior naturally varies, this observation provides a valuable reminder about diversity in language learning. Indeed, a value of any longitudinal study is the ability to measure and describe change in outcomes at the individual level.

Previous research has indicated that the increase in metaphor density of open-class words grows more quickly than that of closed-class words. Littlemore et al. (2014: 128) suggest that

the proficiency level at which this happens may be meaningful, possibly indicating "a qualitative type of metaphor that learners need to use" at higher levels (i.e. B2 and above). To investigate whether such a trend also holds true for the data in the present study, Figure 4 presents a panel of individual scatterplots for each pupil where metaphor density per text *and individual* is plotted as a function of time, measured in semesters. Each plot contains two trendlines per pupil: the dashed grey lines indicate the variation in metaphor density per semester for open-class words, whereas the solid black lines indicate the variation for closed-class words.

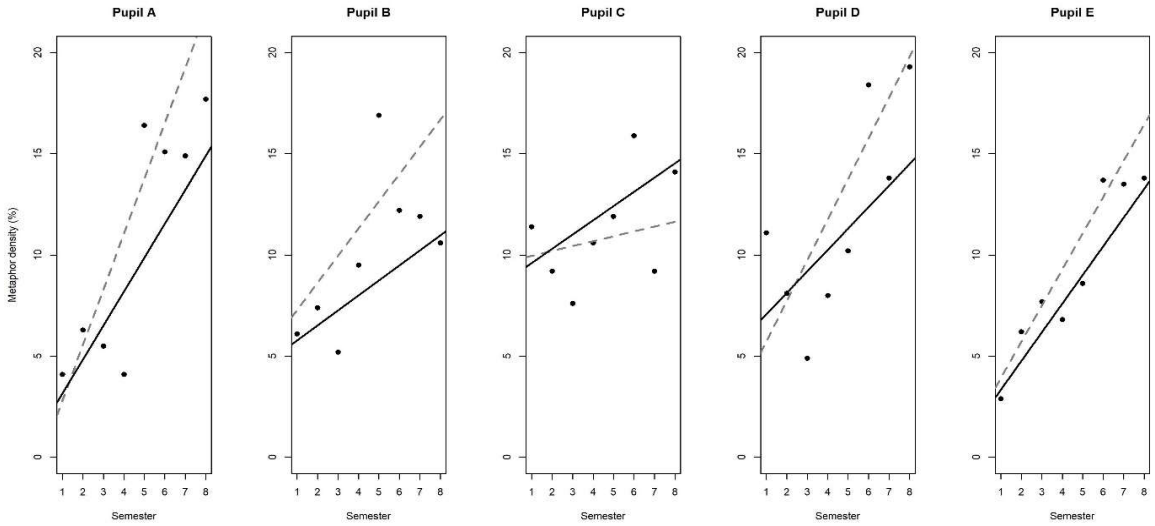

Figure 4: Metaphor density per semester for text and pupil: open-class and closed-class words.

Here we see that the metaphor density of open-class words increases more rapidly than that of closed-class words for four of the five pupils (that is, the trend slopes for the former are steeper than those for the latter), whereas the texts written by Pupil C show the opposite pattern. A binomial test shows that the null hypothesis that there is no difference between metaphor density of open-class and closed-class words over time cannot be rejected (p>0.5). Moreover, the difference in the rate of increase between metaphor density of open-class and closed-class words for the five pupils is not statistically significant (t=1.8618, df=4, p=0.136). What this means is that previous findings regarding the differences between the behavior of open- and closed-classed metaphors are not supported by the data here.

Further, it should be noted any correlation between metaphor density and semester is not an unambiguously causal relationship. More specifically, variations in topic may play a confounding role, affecting metaphorical production. Data from the TRAWL corpus, consisting of texts that are naturally produced in the English language classroom, mirrors a general progression in

language learning. Younger and less proficient pupils are typically asked to produce texts about concrete subjects, where metaphor plays a less central role. As discussed in section 3.1, pupils particularly in grades 8 and 9 were prompted to produce narrative stories or descriptive texts focusing on concrete actions and entities. In the later grades, topics gradually changed to more complex issues. In general, more abstract topics have a greater potential to trigger increased use of metaphor (see also Nacey, 2019). While conformity of topic is a challenge for metaphor research generally (see e.g. Golden, 2012), it may be especially difficult to ensure in longitudinal studies, particularly when researchers are unable to influence topic choice. That said, even if the pupils had responded to identical triggers each semester, such repetition might also have affected their written production; task familiarity has been demonstrated to play a role in language production (see Aas & Nacey, 2019, p. 44).

4.2 Does the distribution of metaphor clusters vary across grade levels and individual?

The scatterplot in Figure 5 shows the observed occurrence of metaphor clusters per text and semester.

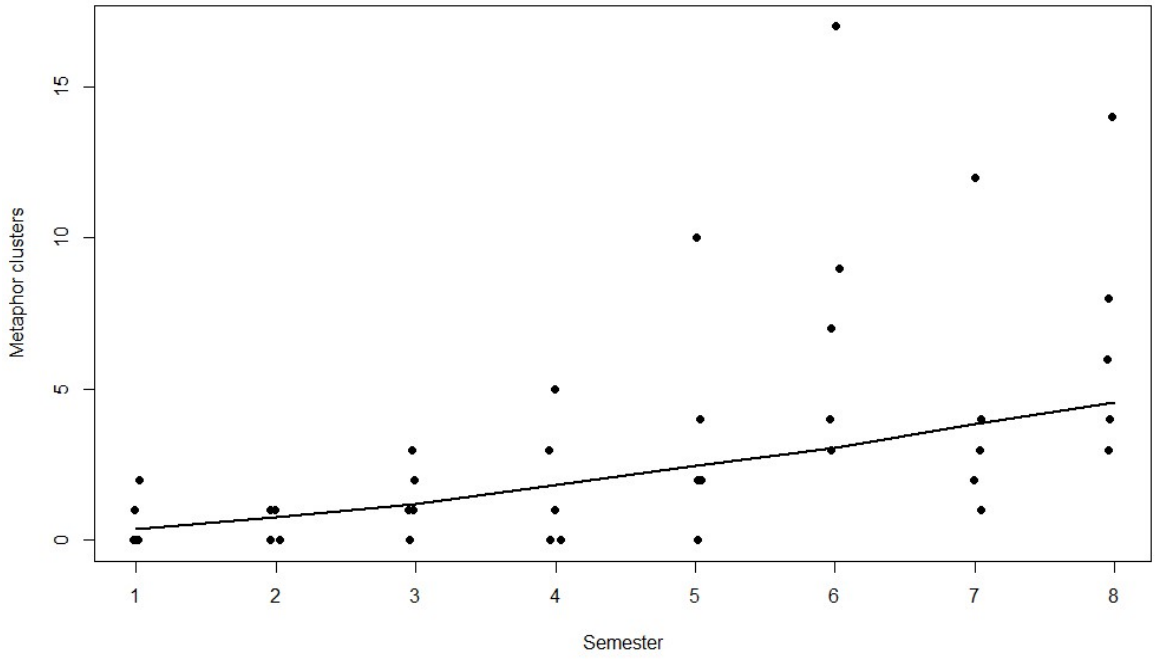

Figure 5: Metaphor clusters per text and semester.

Although metaphor clusters were produced in all grades, we find many texts in grades 8 and 9 (corresponding to semesters 1 to 4) that contain no clusters at all. This testifies to the relatively low level of metaphor in texts written by the youngest pupils in this study. This may, in turn,

reflect their proficiency level and/or result from the concrete nature of the topics the pupils are asked to write in these grades. By the spring semester of grade 10 (semester 6), all pupils produced clusters in their texts, including one outlier with 17 clusters — the maximum number of clusters in a single text in the present data. This tendency continues through grade 11, where all texts contain metaphor clusters, mirroring the trend uncovered in Nacey's (2019) pseudo-longitudinal investigation of TRAWL texts.

While the trendline in Figure 5 indicating a slow, yet steady increase holds true for the group on average, the individual portrayals of the trends for all five pupils show — not surprisingly — some variation in production of metaphor clusters: see Figure 6.

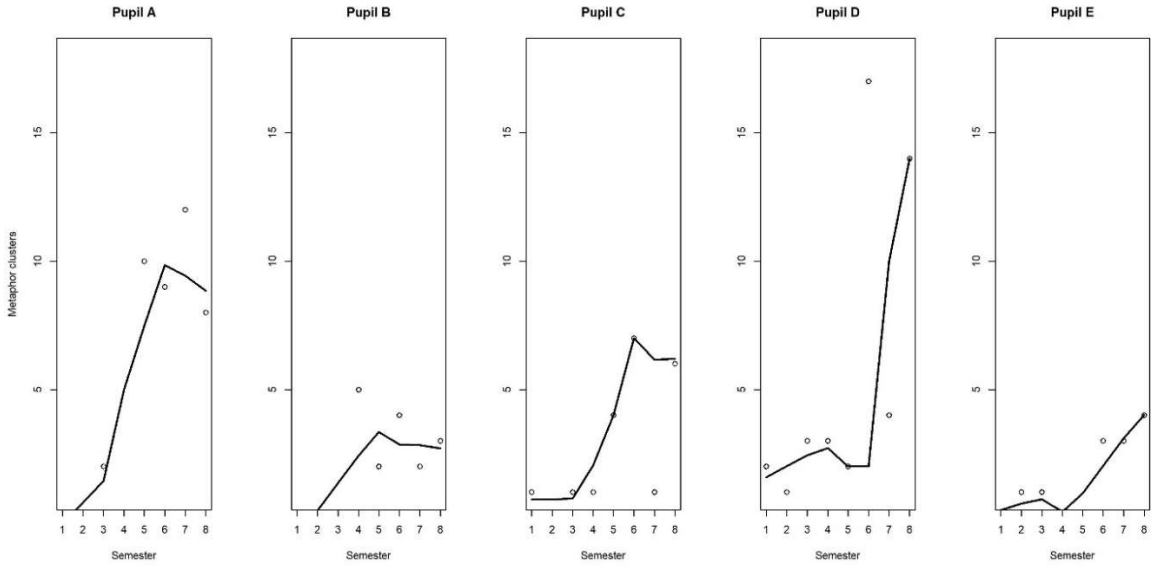

Figure 6: Metaphor clusters per pupil and text, across semesters.

Nevertheless, the occurrence of metaphor clusters becomes more prevalent for *all* pupils in the later grades, even though such progress is not necessarily linear. Pupil B, for example, produced a spike of five clusters in semester 4, followed by a drop to three clusters in semester 5. Even so, there is a clear difference in the nature of these particular clusters: whereas this pupil's semester 4 clusters consist of brief and sometimes overlapping spans, one of the pupil's three semester 5 clusters consists of a long metaphorical analogy (reaching 70% metaphor density, the maximum value in the data as a whole).

4.3 What is the function of metaphor clusters in L2 learner language?

Metaphor density charts detect changes in the rate of metaphoricity over time, but they are agnostic concerning the purpose of any identified clusters. This section therefore explores the metaphor clusters in more depth, focusing on their various functions in the texts.

4.3.1 Function of metaphor clusters in grades 8 and 9

When it comes to the metaphor clusters alone, there seems to be little discernible motivation for them in the texts from these earliest grades. Consider 0, for example, from a ninth-grade essay requiring "a creative text about a person who has a hobby that has played an important part in their life".

> (8) I like volleyball, because it is a team sport, / *where* you *meet* people *with* the same interest as **yourself**. The best *thing about* volleyball is that you play *in* / teams. (4C)

Unlike the cluster in 0, this cluster is single-pointed, having reached the 30% cut-off for what is considered a metaphor cluster by virtue of six words that — while all metaphorical — do not play any coherent, definable role as a whole. This type of cluster lends support to Corts' (2006) observation that some clusters may be random, simply the result of coincidental location within the same word span.

That said, while most clusters from grades 8 and 9 in the data are arguably no more than random, there are nevertheless two clusters that seem to play more of a defined role. We see such potential in 0, from Pupil C's eighth grade spring semester exam, where the selected task involved writing "a letter to your head of school asking him to consider 4 points your group thinks are important when rebuilding the school."

> (9) If / you do al of *this thing* for the pupils, I **think pupils, will *have*** fun at school, and *look forward to* the school. / (1C)

Being one of the concluding lines of the pupil's letter, the sentence appears at a key point in the text. The persuasive function of this cluster may thus be intensified, due to its end location. In brief, this may be an early example where a metaphor cluster plays a discoursal, intensifying function. In grade 9, by contrast, we find a cluster that plays a topical role: see 0.

(10) The / sound is pretty *similar to* when the *air* is *slowly* ***flowing out*** **of a** ***balloon*****. One is** down, two to *go*. Now it is the two / guys standing in front of the main door's turn. (3A)

Here we find a direct metaphor marked by the metaphor flag *similar to*, where Pupil A likens the sound of a trigger to that of the hiss of a leaky balloon. Such an explanatory function is naturally suited to metaphor, allowing us to shed light on a less familiar entity or concept through one that is more familiar.

4.3.2 Function of metaphor clusters in grades 10 and 11

More, and often longer, topical clusters may be identified in texts from the two later grades in the present data. An example is found in Text 5B, whose moving metaphor density chart is presented in Figure 7. Here we find two clusters, the first of which was cited as extract 0, and serves as an introduction to the 103-word long sequence forming Cluster 2.

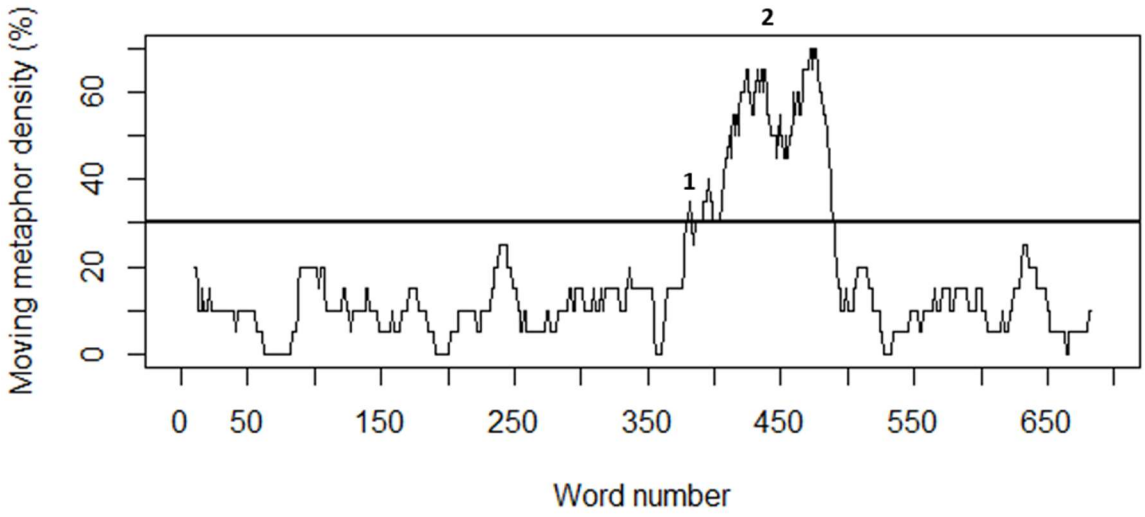

Figure 7: Moving metaphor density chart, Text 5B.

As explained in section 4.1, the exam question asked the pupil to reflect on the sibling relationship in a short story dealing with the topic of forced marriage. In Cluster 2, the pupil produces a long analogy comparing the phenomenon of forced marriage with that of the transatlantic slave trade, including a number of facts about the slave trade from authoritative sources and closing the analogy (and the cluster) by explaining "This is, as told in the start of the example, a very extreme example." The pupil is thus taking strategic advantage of the power of explanatory analogy. As noted in section 3.1, the English exams are intended to test not only

language ability, but also subject knowledge. A sudden shift in topic, which is the hallmark of metaphorical analogies, therefore allows this pupil to demonstrate his/her mastery of subject knowledge beyond what might otherwise be possible, given the essay topic. Note that similes and analogies such as those in extract 0 and in this text are naturally identified as clusters when following the MIPVU protocol for metaphor identification, because all the lexical words in such direct metaphors are coded as metaphorical words (see section 3.2).

Whereas Cluster 2 in Text 5B consists of a single dense metaphor cluster, a different pattern is seen in Figure 8. This figure shows the moving metaphor density chart for Text 6C, an exam response about how the English language connects people: "Create a text in which you reflect on your own experiences of using and learning English when connecting online and offline."

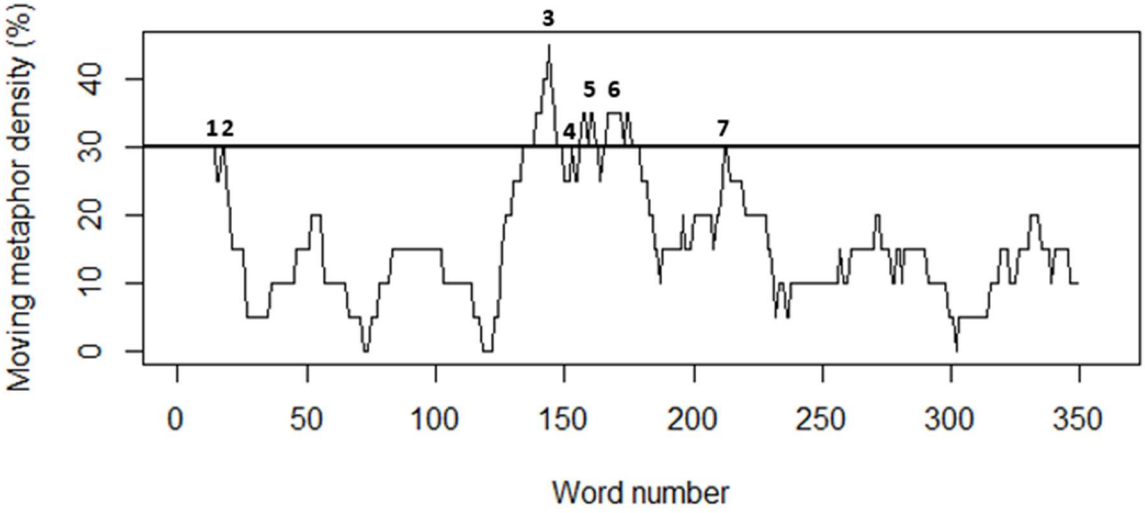

Figure 8: Moving metaphor density chart, Text 6C.

Seven metaphor clusters have been identified in this text, with a particularly dense patch of four consecutive clusters, beginning at word 124 and ending at word 188. Although separated by brief spans dipping below the 30% cut-off, these four clusters are nevertheless interlinked, with overlapping 20-word spans centered in a single paragraph of the text: see extract 0. Here, the bold numbers indicate the start and end of the stretches of text corresponding to Clusters 3-6 depicted in Figure 8. While the metaphorical words are italicized, the bold font indicates the center of the entire series of clusters. More specifically, the four words *the online lives of* are captured in all four clusters, located at the start of Cluster 6, midway in Clusters 4 and 5, and at the tail end of Cluster 3.

(12) The computer games that appear to be most effective for the development of English vocabulary **3/** are those known as Massively multiplayer online role-playing games. *Build* friendship *by connecting to* the digital *world*.
4/ A new report *on* "Teens, **5/** Technology and friendships" *puts* an unusually *positive spotlight on* **6/ the online lives of /3** American teenagers as their **/4** *build* friendship and *connections in* a digital *world.* Teenagers are **/5** *finding* another *ways* to *strengthen* their relationships *with* real-world friends as well as *making* new friends *through* social **/6** media and video gaming. (6C)

This text expands upon the 'connecting' metaphor from the exam prompt, adding the metaphorical actions of *building, finding, strengthening* and *making*, along with the action of *putting an unusually positive spotlight on* online lives. Although the link between these metaphors is topic-based rather than conceptually based, their interaction arguably adds to the persuasive nature of the argument. Cameron and Low (2004) find similar examples in informal L1 English academic writing, attributing it to what they call 'attraction'. By this, they mean that such clustering may result from improved efficiency of the text producer: "having activated a particular conceptual domain with the first use, it is more efficient to remain in the same, closely linked, domains when subsequent lexical choices are made" (p. 368). The present study indicates that attraction may also play a role in the production of metaphor clusters in the L2 English of relatively young pupils.

A discourse management function of metaphor clusters begins to play a more prominent role in grades 10 and 11, with clusters evident especially at the openings and conclusions of the texts. Nine of the texts from these grades begin with metaphor clusters and eleven conclude with clusters, whereas no text in grades 8 or 9 either starts or ends with them. Extract 0 presents Cluster 1 in Text 6A, whose metaphor density chart was portrayed in Figure 2 (section 3.3). The text is entitled *The key to learning English?* and this pupil also responded to the question about how English connects people in today's globalized world.

(13) English is the most important language that / exists. It is the most spoken language worldwide, *after* Mandarin. ***This* is a *great thing*, because it *connects* us and *makes* us able to** communicate *in* a *way where* we all understand what / we are trying to convey. (6A)

This is one of the few clusters in the entire material that reaches 40% density, and is arguably a means of starting the text with a 'grand' vision, before the student continues by recounting how his/her personal experiences with English online and offline contributed towards solving what the concluding metaphor cluster calls *the extremely complex puzzle of learning a language*. Together, these starting and ending clusters serve to frame the pupil's main examples and arguments.

Even though there are more metaphor clusters in the upper grades than in the lower grades that seem to serve some distinct topical or discoursal function, there are nevertheless some in the upper grades that do not appear to be motivated by any defined textual function. For instance, consider extract 0, one of six clusters in Text 8C, where the topic concerns a comparison of the experiences of two characters chosen from literary texts or films that have been studied.

> (14) I personally think the biggest reason he changed was / the challenge *with* the baby. *In* fact, *take* care of **the baby *turn* his mindset** *to* more decent. When it *comes to* Gerry, I / personally think... (8C)

Even though this cluster contains several metaphors from the open-word classes rather than just closed-class metaphors, there nevertheless appears to be no particular function that this concentration of metaphor contributes to the text, other than voicing a complex idea.

Indeed, statistical probability dictates that a number of metaphor clusters are likely to appear by chance, given a certain text length, metaphor density, and cut-off for what is considered a cluster. The histogram in Figure 9 shows the range of possible outcomes (i.e. numbers of metaphor clusters at a 30% cut-off rate) and the probabilities they will occur given a particular set of values: in this case, a text that is 559 words long with a metaphor density of 14.1%. These values match the characteristics of Text 8C.

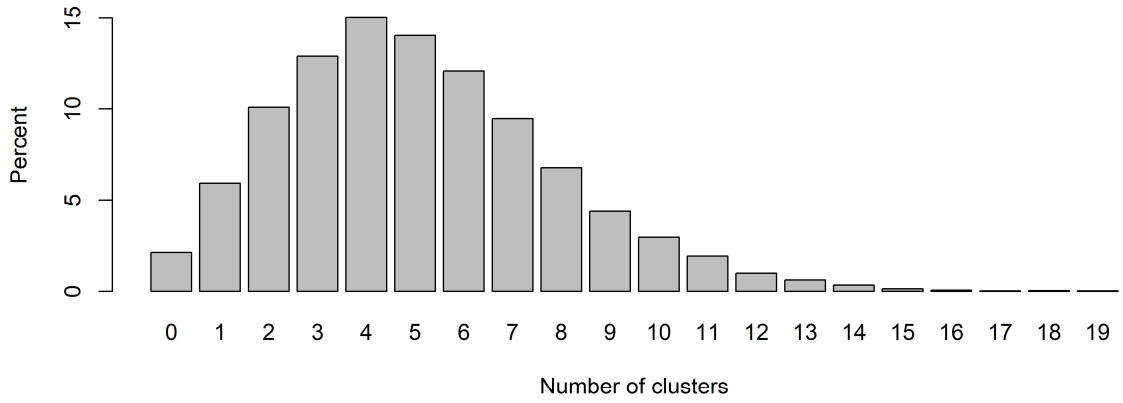

Figure 9: Monte Carlo simulation for Text 8C conditions.

This 'Monte Carlo' simulation calculates results 10,000 times and assumes that the occurrence of a metaphorical word is totally unrelated to the occurrence of a metaphor before or after.[7] Here we see, for instance, that there is a roughly 12% chance that six clusters will be randomly generated given the input variables (1208 out of 10,000 iterations), although the occurrence of four or five clusters is slightly more likely, at approximately 15% and 14% probability respectively. That some metaphor clusters in the data have no discernible discoursal motivation is therefore to be expected.

5. Concluding thoughts

This article has explored the development of metaphor production among young Norwegian L2 learners of English. This investigation builds upon previous research while simultaneously adding a longitudinal perspective that is thus far unique in the field. Because a longitudinal study yields multiple measures for each subject, it allows us to measure changes in outcomes at an individual level, in addition to those at a group level. This particular study tracked the progress of five pupils through grades 8 to 11, from the ages of 13 to 17.

When it comes to quantifiable measures, this study has demonstrated that the metaphor density of texts increases as language proficiency grows. Moreover, the number of observed metaphor clusters, defined here as 30% metaphor density over a 20-word span, also rises steadily over time. Whereas several texts in the earliest grades contain no clusters, they are found in all

[7] Thanks to Dr. Bård Uri Jensen (Inland Norway University of Applied Sciences) for the suggestion of applying a Monte Carlo simulation to this data and for writing the relevant R code (MetaphorClusterSimulation.R), also available as supplemental material at https://dataverse.no/.

texts from the grade 10 spring semester and onwards. These trends are true both generally as well as for each of the five pupils, and agree with the results of earlier studies. However, the present study finds no significant difference in the rates of increase for open-word versus closed-word classes, and is thus at odds with the findings of previous studies in this matter. Littlemore et al. (2014) find that the point at which the rate of increase of metaphor density of open-class words overtakes that of closed-class words correlates with the B2 level of proficiency; that this point does not seem to have been reached in this TRAWL data may simply indicate that these learners have not reached the B2 level, but could equally indicate that Littlemore et al.'s (2014) findings may not be universally applicable and should be tested with additional data. That said, when the sample size is as small as it is here — with only five informants — the power of any statistical test is necessarily weak. The particular strength of the present study is therefore its focus on the individual nature of language learning, providing case studies showing areas in which the behavior of individual pupils both match and deviate from possible general trends.

Regarding the qualitative analysis, this study has explored the possible function of the metaphor clusters identified in the texts, building upon findings from previous studies indicating that clustering has been observed to be meaningful for discourse management or topic. For the present data, however, many metaphor clusters had no such discernible motivation. This was true especially for grades 8 and 9 (the two lowest grades investigated here), although there are a few clusters even at these levels that may arguably play a specific role. Possible functions of metaphor clusters in the two subsequent grades can more often be detected, with clusters frequently appearing at important junctures in texts (particularly the opening and closing paragraphs). The phenomenon of metaphor attraction may also be at work in some clusters, adding to the persuasiveness of an argument. Additionally, a particularly strategic function of metaphor clusters was noted, where a pupil took advantage of the affordances of metaphor to link subject knowledge (about the slave trade) with the topic at hand (the protagonists in a short story). This function has hitherto gone unnoted in the literature about the discourse role of clusters.

Previous literature on metaphor clusters has not detailed 'negative' results. Rather, earlier studies instead focus on clusters whose functions appear readily explainable, thereby giving the impression that all metaphor clusters serve a particular function by virtue of their metaphorical nature. Given the statistical probability of the random generation of a certain number of clusters for *any* text of a given length and metaphor density, however, it seems surprising that so little mention is made in past publications of random clusters. Identification and investigation of metaphor clusters is a valuable tool in discourse analysis, but reports of random 'noise' for

different text types would also be valuable. For instance, a preponderance of seemingly unmotivated clusters might be a hallmark of novice texts, such as those in the present study, where pupils have not yet mastered their writing skills in either their L1 or L2. Text topic may also play a decisive role in the function of metaphor clusters, just as for metaphor density generally.

Not surprisingly, the final word about the topic of the development of L2 metaphorical production has not yet been written. In particular, future investigation into this area should ideally control for the factor of topic uniformity, and data from a greater number of informants should be analyzed. This is a considerable challenge for longitudinal or pseudo-longitudinal research, especially involving young learners as they mature through their school careers, although newly developed corpora offer exciting possibilities in this regard.[8] While the present study thus represents a first look offering a longitudinal perspective on the development of an important facet of L2 metaphorical competence, more such research may soon follow.

Data Availability Statement

Replication and supplemental data that support the findings of this study are openly available in the online data depository Dataverse.no at https://doi.org/10.18710/5C0RSF.

References

Aas, H. L., & Nacey, S. (2019). Methodological concerns for investigating pause behavior in spoken corpora. In L. Degand, G. Gilquin, L. Meurant & A. C. Simon (Eds.), *Fluency and Disfluency across Languages and Language Varieties* (pp. 41–64). Presses universitaires de Louvain.

Cameron, L., & Low, G. (2004). Figurative variation in episodes of education talk and text. *European Journal of English Studies*, *8*(3), 355–373. https://doi.org/10.1080/1382557042000277430

Cameron, L., & Stelma, J. (2004). Metaphor clusters in discourse. *Journal of Applied Linguistics*, *1*(2), 7–36.

Corts, D. P. (2006). Factors characterizing burst of figurative language and gesture in college lectures. *Discourse Studies, 8*(2), 211–233.

Corts, D. P., & Meyers, K. (2002). Conceptual clusters in figurative language production. *Journal of Psycholinguistic Research*, *31*(4), 391–408.

Corts, D. P., & Pollio, H. R. (1999). Spontaneous production of figurative language and gesture in college lectures. *Metaphor and Symbol*, *14*(2), 81–100.

[8] One such corpus is LEONIDE, Longitudinal Learner Corpus in Italiano, Deutsch and English (launched in 2022), where topics were repeated in the first and third year of text collection from young learners: see https://www.porta.eurac.edu/lci/leonide/.

Cuberos, R., Rosado, E., & Perera, J. (2019). Using deliberate metaphor in discourse: Native vs. non-native text production. In I. N. Ferrando (Ed.), *Current Approaches to Metaphor Analysis in Discourse* (pp. 235–256). De Gruyter Mouton.

Deignan, A., Littlemore, J. & Semino, E. (2013). *Figurative language, genre and register*. Cambridge University Press.

Dirdal, H., Hasund, I. K., Drange, E-M., Vold, E., & Berg, E-M. (2022). Design and construction of the Tracking Written Learner Language (TRAWL) corpus: A longitudinal and multilingual young learner corpus. *Nordic Journal of Language Teaching and Learning*, *10*(2), 115–135. https://doi.org/10.46364/njltl.v10i2.1005

Golden, A. (2012). Metaphorical production in L2 writing. In F. MacArthur, J. L. Oncins-Martínez, M. Sánchez-García & A. M. Piquer-Píriz (Eds.), *Metaphor in use: Context, culture, and communication* (pp. 135–173). John Benjamins.

Hoang, H. (2014). Metaphor and second language learning: The state of the field. *TESL-EJ, 18*(2), 1–27. http://tesl-ej.org/pdf/ej70/a5.pdf

Hoang, H., & Boers, F. (2018). Gauging the association of EFL learners' writing proficiency and their use of metaphorical language. *System, 74*, 1–8. https://doi.org/10.1016/j.system.2018.02.004

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. Routledge.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Littlemore, J., Krennmayr, T., Turner, J., & Turner, S. (2014). An investigation in metaphor use at different levels of second language writing. *Applied Linguistics, 35*(2), 117–144.

Littlemore, J., & Low, G. (2006). *Figurative thinking and foreign language learning*. Palgrave Macmillan.

Nacey, S. (2013). *Metaphors in learner English*. John Benjamins.

Nacey, S. (2017). Metaphor comprehension and production in a second language. In E. Semino & Z. Demjén (Eds.), *The Routledge handbook of metaphor and language* (pp. 503–516). Routledge.

Nacey, S. (2019). Development of L2 metaphorical production. In A. M. Piquer-Píriz & R. Alejo-González (Eds.), *Metaphor in foreign language instruction* (pp. 173–198). Mouton de Gruyter.

O'Reilly, D., & Marsen, E. (2021). Eliciting and Measuring L2 Metaphoric Competence: Three Decades on from Low (1988). *Applied Linguistics, 42*(1), 24–59. https://doi.org/10.1093/applin/amz066

Pollio, H. R. (1977). *Psychology and the poetics of growth: Figurative language in psychology, psychotherapy, and education*. L. Erlbaum Associates.

Pollio, H. R., & Barlow, J. M. (1975). A behavioural analysis of figurative language in psychotherapy: One session in a single case study. *Language and Speech*, *18*(3), 236–254.

Rørvik, S., (2022). Noun-phrase complexity in the texts of intermediate-level Norwegian EFL writers: stasis or development? *Nordic Journal of Language Teaching and Learning*, *10*(2), 298–326. https://doi.org/10.46364/njltl.v10i2.987

Semino, E. (2008). *Metaphor in Discourse*. Cambridge University Press.

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. John Benjamins.

Turner, J. (2010). *Investigating figurative language in EFL learners' writing across levels of proficiency.* MA thesis, University of Birmingham, Centre for English Language Studies.

Ørevik, S. (2012). From 'essay' to 'personal text': The role of genre in Norwegian EFL exam papers 1996-2011. *Acta Didactica Norge, 6*(1). https://doi.org/10.5617/adno.1090.