Title of Project:
The Nature of Multi-word Vocabulary among Children with English as a First or Additional Language and its Relationship with Reading Comprehension

Researcher:
Sara A. Smith
University of Oxford
Sara.smith@education.ox.ac.uk

Research Supervisor:
Dr. Victoria A. Murphy

Introduction
Assessment of vocabulary in reading comprehension has tended to focus on measuring the number of words in a text the individual knows (Pearson, Hiebert & Kamil, 2007). However, one area of increasing interest is the role multi-word phrases, such as idioms and collocations, play in language acquisition and processing. There are a variety of definitions of multi-word phrases, one prevailing element is that phrases function similarly to individual lexical items; multiple words express a singular meaning or function (Martinez & Schmitt, 2012; Wray, 2002). Thanks corpus linguistics, it is now well accepted that such items are commonly-occurring and essential (Erman & Warren, 2000; Sinclair, 1991). Multi-word phrases offer a vehicle to express a concept concisely and effectively. Research indicates multi-word phrases may be learned and stored holistically and consequently might be better conceptualized as ‘giant lexical items’ (Nippold, 1998, p.106).

There is evidence that formulaic language is of benefit for language processing speed (Pawley & Syder, 1983; Gibbs et al., 1997). The predictability of these fixed expressions is thought to facilitate reading and reading related skills such as encoding, decoding (Poulsen, 2005, p. 77), fluency (Wray, 2002) and reading speed (Siyanova-Chanturia et al., 2011; Conklin & Schmitt, 2006). However, despite their frequent and facilitative nature, opaque multi-word phrases, such as idioms, have been shown to present a challenge to young learners (Cain, Towse & Knight, 2009) and adult second language (L2) learners (Barfield & Gyllstad, 2009).

While much previous research among young learners has focused on how children develop the ability to comprehend figurative language or approach novel idioms (Cain et al.,2008; Gibbs, 1987, 1991; Levorato & Cacciari, 1992, 1995), at present little is known about the nature of multi-word phrase knowledge among children. Few studies have attempted to measure multi-word vocabulary among children at different ages, perhaps due to a lack of available measures.
This area of inquiry may be of particular import for children who learn English as an Additional Language (EAL) in school, approximately 17.5% of pupils in the UK (NALDIC, 2012).

Previous studies have found that learners with EAL lag behind monolingual English speaking peers in vocabulary and reading comprehension (Bialystok, 2010; Garcia, 1991; Verhoeven, 1990) and it is possible that multi-word phrases present a particular challenge for these learners. For monolinguals, research suggests that vocabulary knowledge at ages 5 and 6 years can be strong predictors of later reading outcomes (Scarborough, 2001; Cunningham & Stanovich, 1997; Roth, Speece and Cooper, 2002) and the larger the child’s vocabulary the faster they will learn subsequent words (Biemiller, 2005; Nation, 2001). Reading, in turn, increases vocabulary (Anderson & Freebody, 1985; Stahl & Nagy, 2005; Taguchi, 1997). Research points to vocabulary as a particular variable that constrains language comprehension for learners with EAL (Garcia, 1991; Verhoeven, 1990); comparisons have shown that the EAL children have smaller English receptive vocabularies than monolinguals (Bialystok, 2010). Attainment studies have shown achievement of learners with the EAL in the UK to be lower than monolinguals (Burgoyne et al., 2009); monolinguals have higher levels of educational achievement overall throughout schooling (DCSF, 2006/2007) and there is a significant attainment gap in A level scores between EAL and monolinguals, controlling for socio-economic status (NALDIC, 2006). Previous assessment of vocabulary in children has often failed to include consider multi-word phrases as vocabulary items, despite argument that these word combinations are not processed or stored as separate words and should be included in measures of vocabulary size (Martinez & Schmitt, 2012, Nippold, 1998; Wray, 2002). It is therefore of interest to explore non-transparent multi-word phrase knowledge among both in monolingual children and children with EAL, as these items may have an impact on and a reciprocal relationship with reading comprehension.

Current Study
The aim of the current research was to explore the nature of verb + object multi-word phrase knowledge among monolingual English speakers and learners with EAL in school years 3, 4 and 5 (corresponding to ages 7/8 years, 8/9 years and 9/10 years, respectively). First, the project developed a semi-productive measure to assess knowledge of multi-word phrases of comparable overall frequency and composed of comparably frequently occurring components. Test items were divided into three categories based on opacity: transparent, semi-transparent and non-transparent. This classification system is adapted from Nesselhauf (2005) and Revier (2009). Classification is based on the semantic properties of the word items in the phrase. If the verb and object are both used with their literal, first dictionary entry meaning, the phrase is ‘transparent’; if the noun is literal but the verb takes on a non-literal meaning, the phrase is ‘semi-transparent’; if neither is literal and the phrase forms a whole meaning that could not be derived from the literal parts, the phrase is ‘non-transparent’ (Revier, 2009).

Test format was adapted for young learners from an existing format, the CONTRIX, a format for testing multi-word phrase knowledge among Danish speaking adult L2 learners (Revier, 2009). A prompt sentence is presented along with a matrix of other context-matched high frequency words. The test taker chooses one word from each column to form a phrase that completes the sentence. Only one viable English phrase can be produced, other lexical items in the matrix
cannot form a phrase. Special attention was paid to avoid making the MPT reliant on reading comprehension, sentence prompts were lexically and structurally simple with only literal language. Content required minimal background knowledge and contained only topics hopefully familiar to the age group such as school, friendships and food. The MPT was designed to be short while still containing enough test items to be valid, so as not to tire young learners.

With this Multi-word Phrase Test (MPT), the study investigated whether children would be able to manifest knowledge of multi-word units of differing transparency by administering the task to 108 monolingual and EAL learners along with a battery of assessments for expressive and receptive single word vocabulary, non-verbal intelligence measure. A further goal of the study was to address the gap in our understanding regarding the role of multi-word knowledge as an element of vocabulary that may contribute to variance in performance on four areas of reading ability (single word reading, reading accuracy, reading rate and reading comprehension) among young learners. 20 monolingual English speaking children and 20 Bengali speaking children with EAL in year 4 were given a battery of assessments, including the MPT, expressive and receptive single word vocabulary measures, the York Assessment of Reading Comprehension (YARC) and a non-verbal intelligence measure to explore how MPT performance might predict variance in reading measure outcomes.

Results
Findings demonstrate that verb + object phrase knowledge can be measured with this learner group. Results showed the reliability and validity of the MPT by evaluating test-retest reliability, internal reliability, concurrent validity and construct validity. The MPT yielded test-retest reliability, internal reliability, and correlated with other measures of vocabulary. Analysis showed that the MPT distinguished adequately between year groups; a one-way ANOVA revealed that for both monolingual learners and those with EAL there was a significant difference in overall MPT score between years 3, 4 and 5. Correlations with other administered vocabulary measures showed a strong relationship between performance on the MPT and performance on measures of other types of vocabulary knowledge, although these relationships differed between the two language background groups. There was a strong effect of transparency; transparent phrase items were most accurately answered by learners, followed by semi-transparent and finally, non-transparent, indicating a substantial difference in the challenge posed by transparent, semi-transparent and non-transparent multi-word phrases. This provides further evidence that opaque phrases present a disproportionate challenge to young learners.

There was a significant difference in transparent and semi-transparent item knowledge for monolingual English speakers between years 3 and 4, and a difference in non-transparent item knowledge between years 4 and 5, with no difference transparent and semi-transparent knowledge. Between years 3 and 4, monolinguals have greater knowledge of transparent and semi-transparent multi-word phrases, without a corresponding increase in their knowledge of non-transparent phrases. By contrast, there were no significant year group differences in transparent item performance found among learners with EAL. Year 5 learners with EAL did, however, perform significantly more accurately on semi-transparent and non-transparent items
than year 3 learners. This more accurate performance on non-transparent items in year 5 might be indicative of a burst in idiomatic or non-literal vocabulary or reflect increased exposure.

A further goal of the study was to identify whether monolinguals and children with EAL would differ on this measure at the three time points. Monolinguals and learners with EAL did not differ significantly in year 3 on any of the administered assessments. Among year 4 children, the two language groups differed in all transparency areas; in year 5, monolingual English speakers performed significantly more accurately in all categories except for semi-transparent items. These significant differences in multi-word phrase knowledge are in accordance with the significant differences between the two language groups on the additional language measures administered, as expected, except in year 3.

The last phase of the current research project explored the relationships between YARC performance and the MPT among year 4 monolingual English speakers and Bengali speaking learners with EAL. Hierarchical regression models controlling for non-verbal IQ, single word expressive and receptive vocabulary and language background status showed that multi-word phrase knowledge clearly has an influence on certain aspects of reading. Performance on the MPT accounted for a significant amount of unique variance in single word reading (33%), reading accuracy (8%), reading rate (37%) and reading comprehension (24%), when controlling. Transparent item scores made a significant contribution to the single word reading model (25%), reading rate (25%) and reading comprehension (10%). Semi-transparent items explained 26% of variance in single word, 11% in accuracy, 18% in reading rate and 17% in reading comprehension. Non-transparent item scores made a significant contribution to the single word reading (30%), accuracy (18%), reading rate (18%) and reading comprehension (18%).

**Conclusion**

Multi-word phrases are a key element of the English lexicon and their role in language development for young learners remains underexplored, despite evidence of their common and challenging presence in English usage and the well-acknowledged role of vocabulary in language and reading. The current study was intended to be exploratory in nature; these results provide insight on the nature of multi-word vocabulary within this age group. The current study provides evidence that this aspect of vocabulary knowledge is distinct from single word vocabulary and has a relationship with reading. Given evidence of a relationship with reading comprehension, though not causal, it may be worth considering explicitly teaching multi-word phrases to young learners or reviewing materials intended to young learners to evaluate the presence of idiomatic or non-transparent multi-word phrases with an eye toward appropriateness and difficulty level. Findings emphasize a need for greater attention on multi-word phrase vocabulary; hopefully, the MPT will provide a tool for future researchers.
References


Savage, R. (2006). Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. *Scientific Studies of Reading, 10*(2), 143-164.


