Linguistic Effects on Anagram Solution:  
The Case of a Transparent Language  

Menelaos E. Sarris1,*, & Chris T. Panagiotakopoulos1  

1 Dept. of Primary Education, University of Patras, Patras, Greece  
*Corresponding author: Dept. of Primary Education, University of Patras, University Campus 26504, Patras, Greece. Tel: 30-2610-997-907. E-mail: m.sarris@upatras.gr  

Received: July 12, 2013 Accepted: July 22, 2013 Online Published: August 4, 2013  
doi:10.5430/wje.v3n4p41 URL: http://dx.doi.org/10.5430/wje.v3n4p41  

Abstract  

Anagram solution tasks have been frequently used to assess word recognition processes and relevant research suggests that anagram solution ability is closely related to reading. Recently, the anagram paradigm was utilized to compare reading performance in the Greek language and was found to share significant positive correlation to reading fluency. The aim of the present study is to explore theoretical views with regard to the linguistic effects on solving anagrams in a transparent language with a simple syllabic structure, using custom made software. Results from 76 children illustrate that anagram solution difficulty is influenced by both syllable complexity and grapheme frequency. These variables also explain much of the variation in terms of the number of moves required for solution and the time spent working on anagrams.  

Keywords: anagram solution; reading fluency; syllabic structure; grapheme frequency; ICT  

1. Introduction  

It is generally acceptable that learning to read constitutes a basic prerequisite for all subsequent literacy development as it fosters access to knowledge and information. In modern societies, reading is considered to be one of the most crucial cognitive skills that supports all later school-based learning. However, reading is not an innate skill (Wolf, 2007). It is a relatively lengthy process that requires mastering a large set of strategies.  

Research on reading development has drawn significant attention over the last four decades and various theoretical models (Frith, 1985; Ehri, 1987; Goswami, 1993; Seymour 1990; Seymour & Evans, 1999) were developed in order to examine the process of reading acquisition. A common feature in various theoretical cognitive models of reading processes is the fact that the initial stages of reading suggest the operation of a “visual analysis system” (Coltheart, 1981; Ellis, 2004). This cognitive mechanism is responsible for converting printed letters to abstract representations as well as for extracting information about where letters are positioned within the words (Pammer, Hansen, Kringelbach, Holliday, Barnes, Hillebrand, Singh, & Cornelissen, 2004, p. 1189). Thus, the visual-orthographic analysis system includes the process of encoding abstract letter identities and the relative position of letters within words (Peressotti & Grainger, 1995). It has been documented that deficits in the visual analysis system are related to different types of peripheral dyslexia, such as visual dyslexia (Cuetos & Ellis, 1999) or letter position dyslexia (Friedmann & Rahamim, 2007). When the visual-analysis system codifies the letter positions and arranges letter strings into common orthographic patterns, the abstract letter forms are subsequently processed in the “visual word form” processor for matching word representations (Galante, Tralli, Zuffi, & Avanzi, 2000; Warrington & Shallice, 1980). This system is “... responsible for recognizing the graphic forms of words and relaying that information on to those regions central to phonological and lexical/semantic processing” (Bolger, Perfetti, & Schneider, 2005, p. 94).  

During a word recognition process, access to the mental lexicon is mediated by the recognition of orthographic units such as letters (Courrieu & Lequeux, 1988). An interesting assumption proposed by Courrieu and Lequeux (1988) asserts the existence of a perceptual mechanism that computes letter order information. Their assumption was based on the fact that despite words being formed by an ordered sequence of letters, readers can solve anagrams. Even though anagrams do not share any orthographic neighbors in the lexicon they are comprised by the same orthographic units (i.e. the same letters), with the exception of the orthotactic violation that disrupts the
morphological structure of the word. In the same study, anagrams were found to produce robust effects on word recognition performance. As Courrieu and Lequeux (1988, p. 32) report, “... position-free letter codes are used for accessing the mental lexicon, while the letter order information is also taken into account, but in a separate way”.

Anagram solving tasks are considered to tap high-level cognitive abilities (Grimes & Mozer, 2001). In an anagram solving task a nonsense string of letters must be rearranged to form a word (e.g. ranamga anagram). The anagram paradigm has been frequently used in cognitive research in a number of studies. It has been utilized to assess word recognition process (Cornelissen, Hansen, Gilchrist, Cormack, Essex, & Frankish, 1998; Pammer et al., 2004), to study the mapping procedure between visual orthographic information and lexical representations during reading (Novick & Sherman, 2004; Witte & Freund, 2001). Anagram tasks are also used to examine memory processes (Bernstein, Rudd, Erdfelder, Godfrey, & Loftus, 2009; Cansino, Ruiz, & Lopez-Alonso, 1999) and problem solving strategies (Novick & Sherman, 2003, 2008).

According to Novick and Sherman (2003), anagram solution is achieved through two distinct processes. The first process produces a “pop-out” solution within about 2 seconds after the presentation of the anagram (see also Henin, Accorsi, Cho, & Tabor, 2009). For these insight-like solutions, solvers bare no conscious awareness of having made incremental progress toward the goal. Pop-out solutions usually occur for well-structured words with common orthographic patterns which are consistent with frequent structural forms in the language (Henin et al., 2009). These structural patterns involve phonological regularity and syllabic patterning (Novick & Sherman, 2008). The second process applies in cases where anagram solvers fail to evoke a lexical response. In that case, they utilize a more time-consuming strategy of a serial process by rearranging the letters until they work out the correct word (Novick & Sherman, 2003).

It is important to note, however, that anagram solution is influenced by different types of information. Performance of poor anagram solvers, for example, is affected by pronounceability (i.e. superficial features), whereas good anagram solvers are affected by “structural features” like the number of syllables (Novick & Sherman, 2008). Equally important aspects of anagram solving, among other factors, are the number of letters that constitute the anagram (Kaplan & Carvellas, 1968), the visual similarities between the anagram and the solution word (i.e. the degree that the anagram violates the morphological structure of the solution word), and the frequency of the solution word in the language (Mayzner & Tresselt, 1958). In fact, Chambers (1979) observed an interference effect in lexical decision tasks when anagrams shared visual similarities to real words. For Deloche and his collaborators (Deloche, Ott, & Tavella, 1995), the interference effect proposed in Chambers’ (1979) study is attributed to specific anagram properties (i.e. visual similarity to real words). They suggest that visual-orthographic similarities between anagrams and real words may affect the operation of the orthographic input lexicon and therefore the word recognition process.

Another key feature of the anagram solving process that has been documented refers to bigram frequencies. Prior work in this field revealed that anagrams derived from target words with high bigram frequencies tend to be easier to solve (Gilhooly & Johnson, 1978). In another study (Mayzner, Tresselt, & Helbock, 1964), it is argued that the important issue is the frequency with which bigrams appear in each position within words. It should be mentioned that bigram frequency measures are also found to affect word recognition (Rice & Robinson, 1975).

According to Mielke and Hume (2001), during word recognition process the initial part of the word is special. Salient information tends to occur at the beginning of words and access to lexical information is achieved on the basis of the left edge of words (Cutler, Hawkins & Gilligan, 1985). In that sense, solving anagrams should be dependent on the features of the initial part of the anagrams. The visual word recognition processes in an anagram solving task may rely primarily on orthographic information (Hargreaves, Pexman, Zdrazilova, & Sargious, 2012; Novick & Sherman, 2008). During anagram unpacking, feedback processes play an important role in both the automated mode of letter-string processing and the analytic strategy (Pammer et al., 2004). As it has been reported, the abstract letter identities seem to correspond to the fundamental perceptual unit of visual word recognition (Pelli, Farell, & Moore, 2003). Hence, an anagram solving task could be conceptualized as a lexical access task (Fink & Weisberg 1981); moreover it could be used as a word identification or reading paradigm (Muncer & Knight, 2011; see also Pammer et al., 2004). As Henin and his collaborators (2009, p. 906) argue, anagram solution is “... an extension of natural word-reading ability, which is highly sensitive to structural knowledge of the language and is generally agreed to involve parallel processing in skilled readers”.

In a recent study (Panagiotakopoulos & Sarris, 2013), performance on anagram solving tasks was found to share significant correlation to reading fluency measures. In fact, children’s performance on anagram solution tasks explained 30% of the variance in reading fluency. These results are consistent with recent findings, where skilled anagram solution is acknowledged as an extension of word-reading ability (Henin et al., 2009). This study reports...
analyses of linguistic factors that affect anagram solution. Our main goal is to examine the effects of syllabic structure and grapheme frequency of target words on an anagram solving task and check for presumable differences on the solving performance of average and below average readers. Thus, we measured time spent and users’ number of moves for solving different nonsense anagrams.

Three main research questions are addressed in this study:

(a) Does the target word’s syllabic structure and grapheme frequency affect anagram solution times? The time-for-solution of a nonsense letter string is used as the measure of the ease or difficulty of solution across different syllabic structures and grapheme frequency.

(b) Does letter order affect anagram difficulty? The difficulty level of an anagram applies to the degree that the letter order is dissimilar to the orthotactic order of the target word. Different parts of the target words are presented in the left edge of the anagram stimuli.

(d) Do syllabic structure and grapheme frequency pose the same difficulties to average and below average readers?

2. The Software Used for the Study

For the needs of this study a custom application was developed using Microsoft Visual Studio, according to the prototyping method (Sommerville, 2004). Working in administration mode, the researcher is able to manage two databases. The first database (main) contains an extended number of words. The second database (custom) has a dynamic content (subset of the main database’s content) that is updated for each experiment and includes the selected words of a specific experiment. Among others, the experimenter is able to define:

• the words of the custom database,
• the segmentation ways of every word,
• the word segments’ display order,
• the words’ display order,
• the available time for finding out the correct word.

The application can be configured to provide help to the user when either the first letter of the correct word, or the last, or even both of them appear. In addition, the application can be configured to provide feedback to the user with sound, animation, and written messages, after the completion of each try (Panagiotakopoulos & Sarris, 2013).

In user-mode, the user can “play” with the application, while several counters record diverse user-specific data, such as the segment being moved, the time spent for the movement of each segment, the total time for the movement of all segments, etc. The user has to move all segments of the anagrammed word and place them with an appropriate order near a fixed line (see Picture 1). The movement of every segment is performed through drag and drop and the user can undo any movement. Picture 1 depicts the application running in user-mode while the user is trying to create the anagrammed word.

The application creates a .txt file for each user’s session, which includes all user actions and time parameters related to them. This file can be imported into a spreadsheet program to analyze the performed user actions in a quantitative and/or qualitative manner.

![](image-url)

**Picture 1:** Actual Screenshot When User is Trying to Create the Anagrammed Word
3. Methodology

3.1 Participants

A sample of 76 Greek-speaking 6th grade students (44 boys and 32 girls) was randomly selected from three elementary schools. The participants’ mean age during testing was 11.61 years. Participants were categorized into two reading fluency groups (see Panagiotakopoulos & Sarris, 2013).

3.2 Experimental Procedure

Subjects were tested in individual sessions by a single experimenter, with each session lasting approximately 15 minutes. All testing trials were conducted in the computer laboratory of each school. A group of four children was used to pilot test the software’s calibration settings, ergonomics, and the presentation format of the anagrams. For the main experimental procedure, a typical desktop personal computer with Microsoft Windows 7 and a LCD/TFT display 17” was used. The main experimental procedure consisted of two trial sessions.

In each trial, participants had to solve 36 anagrams of a target word (i.e. NWIDWO → WINDOW). In the first test stimuli were derived from 2-syllable target words (5-6 letters), while in the second test anagrams were obtained from 3-syllable target words (6-7 letters). Two experimental conditions were used in this study. In the first condition, participants needed to reposition the letters of the anagram to produce the target word. In the second condition, anagrams were formed after repositioning letter strings of the target. Participants had to complete six successive testing sessions for each experimental condition. The total number of rearrangements or movements (either single letters or letter strings), reaction time, the total time consumed, and accuracy scores were calculated. The maximum time allotted for anagram solution was 60 seconds.

The anagrams used in this study were allocated into three presentation categories (i.e. First, Middle, and Last). In the First category the initial part of the target word matched the initial part of the anagram (i.e. σπαθι [sword] → σπ-ι-αθ). In the Middle category the middle part of the target word was located in the initial part of the anagram (i.e. καρεκλα [chair] → εκ-λα-καρ) and in the Last category the last part of the target word was presented in the initial position of the anagram (i.e. νομισμα [coin] → μα-μισ-νο). The 72 target words selected for the study were drawn from a database that includes all words in the 6th grade students’ curricular reading textbooks. The target stimuli were of middle frequency (≈.323) and had the most common syllabic structures in the Greek language. The syllabic structures of the items used in the study were Consonant-Vowel (CV), Consonant-Consonant-Vowel (CCV), Consonant-Vowel-Vowel (CVV) and Consonant-Vowel-Consonant (CVC). Relative frequencies were calculated for all vowel phoneme-to-grapheme mappings in the database and the target words that were selected for the study were labelled as to whether they contain frequent or infrequent phoneme-to-grapheme mappings.

4. Results

An analysis of variance was conducted to evaluate the hypothesis that anagram solving time is affected by the syllabic structure of target words. The mean anagram solving times are presented in Table 1. Results were analyzed using a 4 x 2 Analysis of Variance, with Syllabic Structure (CV, CCV, CVV, CVC) and Group (Average and Below Average readers) as the between-group factors. The Group x Syllabic structure interaction was not significant, F(3,5286) = .43; p > .05. The main effects for Syllabic Structure, F(3,5286) = 48.28; p < .001 and for Group, F(1,5286) = 46.69; p < .001 were found to be significant. The sample means are displayed in Table 1. Post hoc contracts revealed that the average reading fluency group outperformed the below average reading fluency scores in all conditions (p < .001), while significant differences across syllabic structures were also observed.

Table 1: Mean Total Time Scores across Syllabic Structure (Standard Deviations are shown in brackets)

<table>
<thead>
<tr>
<th>Syllabic structure</th>
<th>Average fluency</th>
<th>Below Average fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>10.61 (6.68)</td>
<td>13.96 (8.3)</td>
</tr>
<tr>
<td>CCV</td>
<td>16.12 (16.74)</td>
<td>19.27 (19.71)</td>
</tr>
<tr>
<td>CVV</td>
<td>18.12 (18.22)</td>
<td>20.96 (21.59)</td>
</tr>
<tr>
<td>CVC</td>
<td>13.66 (12.69)</td>
<td>17.86 (19.33)</td>
</tr>
</tbody>
</table>

As Figure 1 depicts, the normal reading fluency group achieved significantly lower solution times for all syllabic structures. Moreover, the simple structures (i.e. CV and CVC) posed the smallest difficulties to participants.
A further analysis was employed in order to check for presumable differences between groups in terms of grapheme frequency. Results were analyzed using a 2 x 2 Analysis of Variance, with Graphemic Frequency (Frequent - Infrequent graphemes) and Group (Average and Below Average) as the between-group factors. The Graphemic Frequency x Group interaction was not significant, F(1,5286) = .02; p > .05. This analysis revealed significant main effects for Graphemic Frequency, F(1,5286) = 161.04; p < .001 and for Group, F(1, 5286) = 39.62; p < .01. The sample means are displayed in Figure 2.

Post hoc contracts found that the average reading fluency group outperformed the below average reading fluency group’s scores in all conditions (p < .001), while significant differences across graphemic frequency were observed with infrequent graphemes being recognized faster (p < .001). It seems that children tend to solve anagrams with infrequent graphemes faster compared to those with frequent graphemes.

A 4 x 2 Analysis of Variance, with Syllabic Structure (CV, CCV, CVV, CVC) and Graphemic Frequency (Frequent - Infrequent graphemes) as the between-group factors was conducted to test the hypothesis that anagram solving time is affected by grapheme frequency. This analysis disclosed a significant Syllabic Structure x Graphemic Frequency interaction, F(3,5286) = 29.11; p < .001. The nature of these interactions is displayed in Figure 3. Subsequent analyses demonstrated that there were simple effects for Syllabic Structure, F(3,5286) = 21.44; p < .001, and for Graphemic Frequency, F(1,5286) = 128.88; p < .001. As Figure 2 shows, participants demonstrated higher total time scores to solve anagrams produced from words that contain frequent grapheme mappings. What is more important is that the effect of syllabic structure is closely associated with the frequency of the containing graphemes. Thus, it seems easier to determine a target word if the nonsense string includes low frequency phoneme mapping of a letter.
Figure 3: Total Time Scores of Anagram Solution for Grapheme Frequency across Syllabic Structure

As mentioned in the introduction, word recognition process is based on the features of the initial part of words. Therefore, we used different parts of the target words as the initial section of the anagrams. The presentation format of the anagrams included three types of stimuli (i.e. the initial, the middle, and the last part of the target word to be presented in the left edge of the anagram). A 4 x 2 x 2 Analysis of Variance, with Syllabic Structure (CV, CC, CVV, CVC), Graphemic Frequency (Frequent - Infrequent graphemes), and Presentation (First, Middle, Last) as the between-group factors was conducted to check for presumable differences in anagram solution times across syllabic structure and presentation format. This analysis revealed a significant Syllabic Structure x Presentation interaction, F(9,5286) = 24.83; p < .001. Figure 4 displays the nature of this interaction.

Figure 4: Total Time Scores of Anagram Solution for Presentation Format across Syllabic Structure

Our data suggest that when the first part of the anagram string is formed by the middle section of the target words, the anagram solution time increases. Target words with CVV and CVC syllabic structures were mostly affected. The CVV structure (i.e. words that contain vowel diagraphs) was also found difficult to solve when the last part of the word was presented first in the anagram string.

As Figure 5 depicts, the graphemic frequency factor seemed to have a significant effect on anagram solving time. In particular, when the middle part of the target word was located at the beginning of anagram strings, participants achieved higher time scores in the frequent graphemic condition.

A Graphemic Frequency x Presentation interaction was also observed, F(3,5286) = 8.62; p < .001 (Figure 5). Subsequent analyses demonstrated that there were simple effects for Syllabic Structure, F(3,5286) = 40.25; p < .001, and for Presentation, F(2,5286) = 180.23; p > .001. As Figure 3 shows, when the middle or last letter of words were presented in the initial position of the anagrams, participants demonstrated higher time scores for the CVV and CVC words. The simple effect for Graphemic Frequency also proved to be significant, F(2, 5286) = 62.85; p < .001.
Prior work in the Greek language suggested that performance on anagram solving tasks is highly correlated to reading fluency (Panagiotakopoulos & Sarris, 2013). In an effort to elaborate further on this issue, we investigated the relationship between average and below average readers, comparing the number of required moves to solve the anagrams and the linguistic factors of syllabic structure and graphemic frequency.

Results were analyzed using a 4 x 2 Analysis of Variance, with Syllabic Structure (CV, CCV, CVV, CVC) and Reading fluency (average - below average) as the between-group factors. This revealed a significant Syllabic Structure x Reading fluency interaction, F(3,3552) = 3.07; p < .05, which is displayed in Figure 6. Subsequent analyses demonstrated that there were simple effects for Syllabic Structure, F(3,3552) = 13.82; p < .001, and for Reading fluency, F(1,3552) = 39.29; p > .001.

A final analysis was carried out in order to search for possible differences between the two reading groups on the graphemic frequency factor. Results were analyzed using a 2 x 2 Analysis of Variance, with Graphemic Frequency (Frequent - Infrequent graphemes) and Reading fluency (average - below average) as the between-group factors. The analysis, as displayed in Figure 7, revealed significant main effects for Graphemic Frequency, F(1,3552) = 15.67; p < .001, and for Group, F(1, 3552) = 37.37; p < .001.

---

**Figure 5:** Total Time Scores of Anagram Solution for Grapheme Frequency across Presentation Format

**Figure 6:** Total Moves of Anagram Solution for Reading Fluency across Syllabic Structure
5. Discussion

The anagram paradigm has recently attracted the attention of researchers (among others, Cornelissen et al., 1998; Grimes & Mozer, 2001; Novick & Sherman, 2003, 2004, 2008; Pammer et al., 2004). Two present the main trends on anagram research. The first one focuses on strategies (Novick & Sherman, 2003, 2008) that anagram solvers employ and examines the linguistic factors that influence anagram solving (Deloche et al., 1995; Novick & Sherman, 2008). Another line of research centers on issues that concern the relationship between anagram solving performance and reading process (Deloche et al., 1995; Henin et al., 2009).

The present study attempts to extend previous research in the field. In Deloche et al. (1995), for example, Italian speaking children were found to share the direct relationship between reading and performance on anagram solving tasks. The authors linked their findings to the transparency of the Italian language as they failed to replicate the results with the French-speaking group. The connection between reading performance in an alphabetic language with simple sound-to-print correspondences and anagram solving abilities has been recently examined in Greek (Panagiotakopoulos & Sarris, 2013). Results revealed that performance on anagram solving tasks is positively correlated to the reading fluency measure used in the study. Following-up the data of the previous study, we proceed to a more detailed analysis in an effort to investigate the linguistic features that affect anagram solution in the Greek language and examine whether some of those features are linked to performance variation among average and below average readers.

Data analysis disclosed that anagram solving time is affected by the syllabic structure of target words. The main effect of syllabic complexity indicates that both groups (i.e. average and below average readers) were equally affected. Significant differences across syllabic structures were also observed, with the simple CV and CVC structures being easier to solve. The effect of syllabic complexity was also noticed in the reading fluency measure, where the average group clearly outperformed the below average readers. Both experimental groups achieved lower time scores at the simple syllabic structures (i.e. CV and CVC).

An unexpected finding, which needs further confirmation, was that both average and below average readers recognized anagrams containing infrequent graphemes faster compared to anagrams with frequent phoneme-to-grapheme mappings. More importantly, the effect of syllabic complexity seems to weaken when grapheme frequency is taken into account. It should be noted at this point that relevant research in deep orthographic writing systems like the English language suggest that anagrams containing bigrams of high frequency tend to be easier to solve. This was not the case in our study, where solvers managed to identify anagrams that included low frequency phoneme-to-grapheme mappings more quickly. A possible explanation could be that, in some cases, anagrams with bigrams that are otherwise infrequent are more sensitive to identification through the visual-orthographic analysis system. In a shallow orthography like the Greek language with simple syllabic structures, perhaps low frequency graphemes facilitate the “pop-out” solution process (Novick & Sherman, 2003) as fewer candidates fulfill the requirements to be the target word. Whatever the case, these results need to be replicated in other studies employing a larger sample.

As discussed in the introduction, word recognition process is based on the features of the initial part of words. Thus, important information on anagram solving could be extracted by comparing solution time scores for anagrams that had different parts of the target words as their initial section (i.e. the initial, the middle, and the last part of the target word to be presented in the left edge of the anagram). The analysis confirmed that the effect of syllabic complexity
depends on the level of the other factors such as grapheme frequency and presentation format. The middle condition (i.e. the first part of the anagram is formed by the middle section of the target word) proved to be the most demanding in terms of solution time scores for both CVV and CVC syllabic structures. Anagrams with vowel diagraphs (CVV structure) were also found to be difficult to solve when the last part of the word was presented first in the anagram string. It should be stressed here that solution time scores were higher when the middle section of the target words was presented at the beginning of the anagrams in the frequent graphemic condition.

In a previous study in the Greek language, reading fluency was found to share strong correlation to performance on anagram solving tasks (Panagiotakopoulos & Sarris, 2013). Following up on this issue, we examined the association between reading fluency and the linguistic factors of syllabic structure and graphemic frequency by comparing the number of moves required to solve the anagrams. The average group solved anagrams using significantly fewer moves than the below average fluency group. The only exception was detected on the simple CV syllabic structure where no difference between groups was traced. The analysis of the graphemic frequency factor duplicated previous results. Anagrams containing frequent graphemes required fewer moves to be solved.

6. Conclusion

In this study, with the use of custom made software, we made an effort to explore theoretical views with regard to the linguistic effects on solving anagrams in a transparent language with a simple syllabic structure. In general, the results showed that anagram solving time is affected by the syllabic structure of target words. The effect of syllabic complexity (that it is dependent on factors such as grapheme frequency and presentation format) was also revealed in the reading fluency measure.

From our point of view, the anagram paradigm, which is mainly used by experimental psychologists, could also be used in educational research. Solving anagram tasks may give insights into how readers extract information from a text and how different orthographic features like the complexity of letter-sound correspondences or syllabic structure affect word recognition process. Therefore, the prospect of utilizing anagram tasks as an alternative tool for measuring reading fluency should be further evaluated.

Finally, a point that needs to be further explored is the correlation between the recognition of the anagrams containing infrequent graphemes and those with frequent phoneme-to-grapheme mappings.

References


