Evaluating Phrase Sets for Use with Text Entry Method Evaluation with Dyslexic Participants

Akiyo Kano
Department of Computing
University of Central Lancashire,
Preston,
PR1 2HE
akano@uclan.ac.uk

ABSTRACT
Much of pen input research is focused on the use of the pen devices for text input, with accuracy and efficiency being of great importance for the user experience. In measuring these variables, most evaluations use short text phrases. In this paper, we consider the need to evaluate phrase sets to be used with dyslexic participants in evaluating pen-input devices, and attempt to evaluate phrase sets that are currently used in text entry method evaluation.
The paper looks at how the usability of text entry is determined and then considers the importance of phrase sets in these evaluations. It then discusses why usability evaluations and text entry evaluations for text input devices should be carried out for dyslexic users in addition to the standard evaluations. It identifies areas in which a dyslexic participant may experience difficulties during text input method evaluation tasks, and those that are specific to evaluation of text input using pen input device and handwriting recognition. It also outlines the affects *"trigger words"* have on dyslexic people, and introduce a new program which analyses a given phrase set for the amount of trigger words in the set, and then compares some example phrase sets.
The paper concludes that trigger words is likely to have an impact on a dyslexic person's ability to enter text. The two phrase sets analysed in this paper, both of which are currently used in text entry method evaluation, have a high proportion of trigger words and that in order to ensure the validity of the evaluations, the occurrence of these trigger words should be reduced if they are to be used with dyslexic participants.

Author Keywords
Dyslexia, Dyspraxia, Dysgraphia, text input method evaluation, pen input methods, phrase sets, trigger words.

INTRODUCTION
In early pen devices, the hope was that the pen could be used in a very natural way with handwriting recognition software providing a robust and reliable user experience. It is now acknowledged that handwriting recognition is sometimes unreliable, and many applications for pen devices have moved towards soft keyboards for accurate text input or to digital ink storage for natural handwriting (MacKenzie and Soukoreff, 2002). The efficiency and effectiveness of pen devices for text input continues to be of significant importance to the research community. The following section of this paper introduces the primary research on text input and describes the usual methods.

Text Entry Methods
Text entry research has being going on for a significant length of time, with its first surge of interest in 1970s and early 1980s, and a second wave with the pen based computing in early 1990s (MacKenzie, 2002). There is now a wide range of text input methods available for almost every user need. MacKenzie & Soukoreff (2002) separate input methods into two main categories, keyboard-based and pen-based. They justify this classification by outlining that the key feature of a keyboard-based input method is that, it directly produces machine-readable text (i.e., ASCII characters), which can be indexed and searched. In this classification, pen-based technology requires recognition technology to recognize what the user input, so the data can be converted into machine-readable text. When the pen is acting as a pointer on a soft keyboard, it would be considered as keyboard entry using the MacKenzie & Soukoreff (2002) classification method. Irrespective of how the device is categorised, the ultimate goal of any text input device is to offer a way to input text as accurately and as quickly as possible.

Methods for Evaluating Text Entry
As new text entry methods are devised, there is a need for evaluating how efficient these methods are. Text entry method evaluation originally consisted of typewriters, a stopwatch to measure typing speed and counting errors by hand (Soukoreff & MacKenzie, 2003). Modern research into comparing mobile phone text input methods (Butts and Cockburn, 2001; Silfverberg, MacKenzie & Korhonen, 2000; James & Reischel 2001; Dunlop & Crossan, 2000) and pen-based text entry methods (MacKenzie & Chang, 1999; Chang & MacKenzie, 1994) has lead to a refinement of methods used to evaluate these devices.
A common aim of researches involving text entry is to compare two or more text input methods/devices against each other (MacKenzie & Chang, 1999; Chang & MacKenzie, 1994; Read *et al*, 2001), to see which method or device allows the user to enter text most accurately and efficiently.

In designing these evaluations, MacKenzie & Soukoreff (2003) write:

> *"Among the desirable properties of experimental research are internal validity and external validity. Internal validity is attained if the effects observed are attributed to controlled variables. External validity means the results are generalizable to other subjects and situations"*

This implies that the text entry methods or the devices used become the controlled variable and all other factors should be kept at a constant.

In a typical text entry experiment, the participant is shown a short phrase, and is then asked to enter the phrase into the device while the speed at which it was done and the accuracy at which the text was inputted is measured (Butts and Cockburn, 2001; Soukoreff & MacKenzie, 2004). The error rate of the inputted text is then calculated by comparing the intended set of text and what was actually inputted by the user. When measuring these variables the text shown to the participant itself becomes an independent variable and careful considerations must be made to ensure that it does not cause any variation in the measurements.

Phrase Sets

There have been several input texts suggested for text entry research. One is to allow the participant to enter "*whatever comes to mind*'. This seems to increase the external validity in that the text used is natural to the user. However, as MacKenzie and Soukoreff (2003) indicate, it is impossible to measure the accuracy as we do not know what the participant *intended* to type. Therefore, it is preferred, in most research experiments, to present the participant with a carefully chosen set of phrases.

Several different strategies have been taken in creating phrase sets, some use newspaper sentences and sentences that are suppose to emulate a conversation, (James & Reischel, 2001), others use input phrases that are deemed most natural to the user (Soukoreff & MacKenzie, 2003). All of these phrase sets have been designed to have phrases that are moderate in length, easy to remember and representative of the target language. However, the language that is chosen as representative tends to be chosen with the device that is being evaluated in mind (e.g. deciding on the target language according to whether it is a mobile phone with numeric keyboard or a QWERTY keyboard) and not focusing on who the user is and what the user can do. The participants are assumed to have no language or learning difficulties (or such an issue has not been considered!).

DYSLEXIA

Motivation

One group of users for whom language based tasks are problematic is dyslexics. There have been well documented efforts made in adapting technology for dyslexic users (Keates, 2000; Newell, Booth & Beatie, 1991; Gregor *et al*, 2003), which all recognise that dyslexic users have a different set of requirements and usability issues from non-dyslexic users.

Research shows that 90 – 98% of children with learning disabilities experience fine motor skill or handwriting problems (McHale & Cermak, 1992) and the overall number of children suffering from handwriting problems have been estimated from 5% to 25% of the population (Jongmans *et al.*, 2003). Although there are little research carried out in estimating the number of adult dyslexics, the British Dyslexia Association estimates that there are around 2 million dyslexics in the UK, both children and adult (Ott, 1997). For these people, technology often plays an important role in coping with their Dyslexia. It is often the case that professionals suggest technology as an aid to coping with Dyslexia, and in fact, many dyslexic students in higher education are often granted an allowance for buying computers and portable devices to aid their studies. For example, using technology enables students to spell check their work, present a neater essay, keep their schedule and remind them of things they might forget. However, in order to carry out all these tasks, the dyslexic user must be able to use the technology involved and specifically, to be able to input data. Thus it is important that whatever text entry method is used, it is suitable for dyslexic users.

Dyslexia

Although the exact definition of Dyslexia is yet to be agreed upon by the different academia, it is commonly understood as a condition that affects the person's reading and writing as Waites (1968) defines:

> "*A disorder in children who, despite conventional classroom experience, fail to attain the language skills of reading, writing and spelling commensurate with their intellectual abilities"*.

The British Dyslexia Association (BDA) provides this definition of Dyslexia:

> "*May be described as a specific difficulty in reading; spelling and written language, one or more of these areas may be affected. Innumeracy, notational skills (music), motor function and organizational skills may also be involved, However, it is particularly related to mastering written language although oral language may be affected to some degree"* (Crisfield, 1996).

From these definitions, you could summarise that, dyslexic users mainly suffer from problems related to reading, writing and language in general. However, we must understand that Dyslexia is a "*syndrome with a wide range of symptoms*" (Ott, 1997) and every dyslexic person has his or her own unique range of problems they suffer from. One dyslexic person may have difficulty in carrying out simple sums, whereas another may be brilliant at maths, capable of carrying out the most complex calculations in their head.

When taking part in a text entry method evaluation, the first problem occurs when the dyslexic participant is required to read the phrase shown to them to input. It is often the case that the given phrase is either read incorrectly or the phrase is distorted during the transition of the text from being read to being entered. This problem is shared across any input methods.

Dyspraxia and Dysgraphia

When a pen is being used as an input method, there are further conditions associated with Dyslexia that may affect the user's ability to enter text. *Dyspraxia* is a symptom, which is defined as problems in skilled movements (motor skills) despite abilities within the average range (Dawdy, 1981; Miller, 1986). This affects the sufferer's fine and/or

gross motor skills. Pen input methods, like most other input methods, requires fine motor skills to manipulate the device. Handwriting recognition requires the user to enter letters (which is a set of finely coordinated movement) to be clear enough for recognition. Even soft keyboard input method requires the user to point the pen very precisely.

While Dyspraxia can affect the user's ability to enter text across multitude of input devices (such as keyboards as well as pen inputs), *Dysgraphia* poses a unique challenge to pen input alone, and especially, the use of handwriting recognition with pen input. Dysgraphia is another symptom associated with Dyslexia and is loosely defined in literature as handwriting problems, where the handwriting is distorted or incorrect (Miozzo & Bastiani, 2002; Jongmans *et al.*, 2003). It should be noted that, although it is often the case that a person suffering from Dyspraxia and/or Dysgraphia also suffers from Dyslexia, it is possible for one to be dyspraxic and/or dysgraphic without being dyslexic. Dysgraphia gives rise to a second problem in evaluating pen input devices. After the participant has read the give phrase (correctly or incorrectly), they have to write the phrase by hand, using the device. It is easy to see that the user's inability to produce legible texts will ultimately lower the recognition rate of their handwriting. This is another factor that would not affect an non-dyslexic participant in using a pen input device.

Trigger Words

Dyslexics often encounter words that cause them problems in everyday English language, and thus special considerations must be made in order to maintain the internal validity of any experiment. Davis & Braun (1997) introduce the idea of *Trigger words*, these being words that have abstract meanings, and often have a number of different meanings. There are 217 of these words identified by Davis & Braun, and some examples include 'and', 'the', 'to' and 'it' (A full list can be found on P.234 of Davis & Braun, 1997). These words trip up dyslexics because they do not represent objects or actions and cause disorientation to be triggered, which alters most of the person's senses, including vision and touch. This means that when the participant encounters a trigger word, his or her ability in accurately inputting the text becomes compromised and thus weakens the internal validity of the evaluation.

It is common for a dyslexic person to avoid using words that cause them trouble, and so reducing the number of trigger words would mean that the phrases will be closer to the participants' natural languages, therefore strengthening the external validity of the evaluation.

The first five phrases taken from the MacKenzie & Soukoreff (2003) phrase set "phrases2.txt" (from http://www.yotku.ca/mack/ PhraseSets.zip) are shown below, with the trigger words highlighted:

My watch fell in the water
Prevailing wind from the east
Never too rich and never too thin
Breathing is difficult
I can see the rings on Saturn

It is clear from the above that typical phrase sets are riddled with trigger words. There are 15 trigger words out of 28 words in total in these five phrases, which is a high proportion.

In a text input evaluation, trigger words first cause a dyslexic participant problem when they are required to read a given phrase. As Davis & Braun (1997, Ch.4) explains, when a dyslexic participant encounters a trigger word, it is likely to confuse them. Although with increased concentration, they can continue to read on, repeated exposure to more trigger words will result in the participant becoming disorientated. It is highly likely that, at this stage, if not before, the person will read the phrase incorrectly, perhaps omitting or substituting some of the words. This disorientation will be reinforced each time the participant encounters more trigger word.

This confusion and disorientation will continue when the participant comes to enter the phrase into the device. The participant is already at a disadvantage as they have the incorrect phrase stored in their minds. Due to their disorientation, their ability to tell the body to move in order to input the text, and actually entering text also becomes affected.

It is clear that trigger words have a cascading affect on the participant's ability to perform a text entry task. It also amplifies the difficulties caused by Dysgraphia in writing the phrases down with a pen input device.

DEVELOPING AND EVALUATING A PHRASE SET ANALYSER

This early investigation of trigger words lead to the study that is presented here that took two phrase sets and, by investigating trigger words, examined their suitability for dyslexic users.

The Phrase Sets Analyser

Although the identification of trigger words on few short phrases is easy enough to be done manually, comparing every word in a larger phrase set with a list of 217 words quickly becomes a laborious task with high risk of misidentifying some words. A new program called PHANTIM (Phrase Analyser for Text Input Methods) has been created to identify all the trigger words in a given phrase set, and produce useful statistics about the phrases. In addition to counting the trigger words, and other variables, the program can calculate the percentage of trigger words found in each phrase and the overall percentage of the whole phrase set, whilst also highlighting all the trigger words found in the phrase set. The phrases can then be organized in order of the number of words in each phrase, the number of trigger words in the phrase and the percentage of trigger words in each phrase. Figure 1 shows an example screen shot of the analyser.

The program is at its infant stage and is at the moment only able to analyse a given phrase set and re-organise it in order of given variables, its ultimate aim is to aid the process of creating phrase sets that are suitable to be used with dyslexic participants.

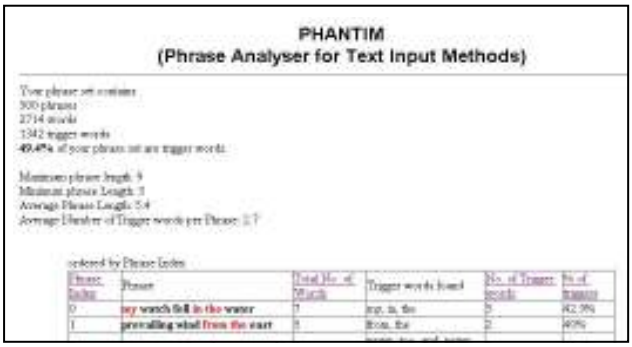


Figure 1: Screen shot of PHANTIM

Analysis of Example Phrase Sets

The two phrase sets created by MacKenzie & Soukoreff (2003) under investigation are; "phrases.txt" which consists of 1004 phrase and "phrases2.txt" which consist of 500 phrases, which have been selected and shortened from the original phrase set. These phrase sets were analysed by PHANTIM. These phrase sets have one phrase per line, with no capital letters for the first letter of the phrase, and little capitalisation elsewhere. They also have no punctuation marks (commas, full stops or apostrophes) and no numbers. The program currently only accepts phrase sets that are in this format. The two phrase sets were fed into the analyzer and the results are shown in the next section.

Results

Table 1 shows the two phrases sets' overall statistics.

	phrases.txt	phrases2.txt
Total number of phrases	1004	500
Total number of words	7642	2713
Total number of Trigger words	3841	1342
Percentage of trigger words in the phrase set	50.3%	49.5%
Maximum number of words in a phrase	13	9
Minimum number of words in a phrase	4	3
Average number of words in a phrase	7.6	5.4
Average number of trigger words per phrase	3.8	2.7

Table 1: Overall statistics of the phrase sets

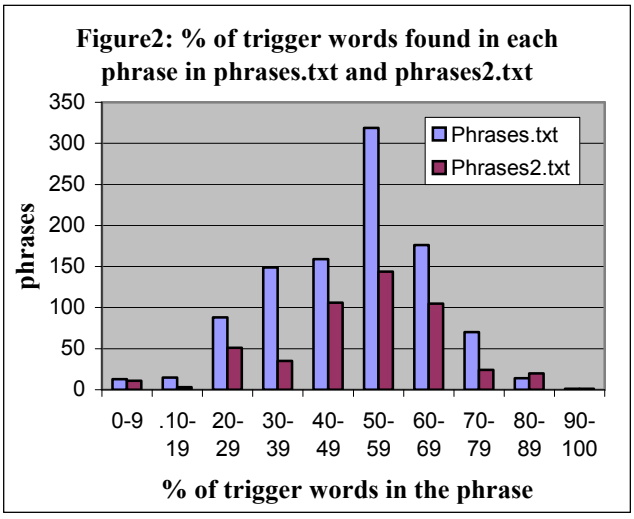
The results show that approximately half of the words in both phrase sets are trigger words. This is a high proportion when you consider that every one of these words causes a dyslexic person to perform less and less well. In comparison, an average of 10 pages from a website for dyslexic readers had a 32% trigger word rate. It should also be noted that some phrases contain higher percentage of trigger words, such as the phrase "`what you see is what you get`", which is 100% trigger words, and some contain no trigger words, such as, "`frequently asked question`".

The percentage of trigger words in each phrase was extracted, and Table 2 shows the percentage of trigger words found in both phrase sets.

% of trigger words in the phrase	phrases.txt	phrases2.txt
0-9	13	11
10-19	15	3
20-29	88	51
30-39	149	35
40-49	159	106
50-59	319	144
60-69	176	105
70-79	70	24
80-89	14	20
90-100	1	1

Table 2: % of trigger words in a phrase

Figure 2 is a bar graph of the data found in Table 2. This shows that in both phrase sets, over half of the phrase contained more than 50% trigger words. The top 50 percentile contains 580 phrases (57.8% of the phrase set) in phrases.txt and 294 phrases (58.8%) in phrases2.txt.



Calculations show that if we take out those phrases containing 50% or more trigger words; the percentage of trigger words in each of the phrase sets decrease to 34% and 32.5% respectively.

Although it is difficult to create a phrase that makes sense without the use of any trigger words, and thus impossible to completely rid the affect on the dependent variable, considerable efforts should be made in reducing the number of trigger words in the phrase set for dyslexics to a minimum.

FUTURE WORK

Currently, PHANTIM can only calculate the statistics of a given phrase set and reorganise them into certain order. Although this has already proven useful in finding out how many trigger words are in the phrase set editing of the phrase set itself is left to the user of the program and they will have to painstakingly copy and paste certain phrases that they want into a text editing software before they are able to use it in text entry method evaluation. In order for the program to aid in the editing of the phrase set, it will

have to be able to gain editing functions. These may include:

- Able to edit the phrase set under certain conditions as "get rid of all phrases containing 50% or higher proportion of trigger words".
- Save and print the edited phrase sets, without all the statistics.

Secondly, the program offers no assistance in *creating* a phrase set that is more suitable for dyslexic participants. It only aids the user to eliminate certain phrases, but then requires them to create new phrases, then test their trigger word rates again. The program should therefore, be able to take in text from books or websites, and create phrases out of the text, where the user can indicate what percentage of trigger words will be tolerated.

There are, so far, no studies in quantifying the effects of trigger words on dyslexic readers. However, in order for text entry method evaluations to have internal validity, investigations should be made on what will be an appropriate amount of trigger words in the phrase set for dyslexic participants. That is, we need to find out at what percentage of trigger words in a phrase set, a dyslexic person is able to perform at the same accuracy in entering text as a non-dyslexic person.

Although there is a clear path in PHANTIM's development as regards to trigger words, this is of course, only one issue out of many other factors that may have an effect on the ability of the participant to enter text. Another issue may be that of the participant's reading age. When a participant encounters a word they have never heard of, their ability to write the word down is lowered and thus reduces the internal validity of the evaluation. A dyslexic person, by definition has a considerably lower reading age than their physical age. It is therefore suggested that the reading age of the participants should be taken into account in the development of an appropriate phrase set.

CONCLUSION

This analysis shows that there is a high percentage of trigger words in the current phrase sets that are used in text entry method evaluation.

In this paper, the need for re-evaluating text entry methods for dyslexic users and the potential problems they can encounter was discussed. It also outlined the need for a new or modified phrase set as an addition to the current popular method of evaluating text entry methods.

By looking at what a phrase set aims to do within an evaluation process, which is to provide the highest internal and external validity, it is clear that trigger words, which cause a dyslexic person to become confused and not able to perform as normal, influence the text he or she inputs into a device and therefore reduces internal validity. For the evaluation process to have high internal validity, the phrase set must have as few of these trigger words as possible.

PHANTIM was introduced as a tool in analysing phrase sets for the amount and density of trigger words in phrase sets. We ran an analysis of two phrase sets, which are currently used in text entry method evaluation and found that approximately half of the words in the phrase sets were trigger words. However, we are as yet unsure about how much reduction of the occurrence of trigger words is required, and thus, further studies are needed. Analysing a phrase set for the amount of trigger words it contains is only the first step in creating a phrase set that is suitable for use with dyslexic participants for pen input devices.

REFERENCES

Butts, L. & Cockburn, A. (2001), An Evaluation of Mobile Phone Text Input Methods, *Third Australasian User Interfaces Conference,* Melbourne, Australia, Conferences in Research and Practise in Information Technology, Vol. 7, pp.55-59.

Chang, L., & MacKenzie, I. S., (1994), A Comparison of Two Handwriting Recognisers for Pen Based Computers, *Proceedings of CASCON '94*, pp. 3364-371. Toronto: IBM Canada.

Crisfield J. (ed), (1996), *The Dyslexia Handbook,* BDA, Reading.

Davis, R. D., & Braun, E. M., (1997), *The Gift of Dyslexia*, Souvenir Press Ltd.

Dawdy, S. C., (1981), Paediatric Neuropsychology: Caring for the Developmentally Dyspraxic Child, *Clinical Neuropsychology*, Vol. 8, pp.367-79.

Dunlop, M. D., & Crossan, A., (2000), Predictive Text Entry Methods for Mobile Phones, *Personal Technologies*, pp.134-143.

Gregor, P., Dickinson, A., Macaffer, A., & Andreasen, P., (2003), SeeWord – a Personal Word Processing Environment for Dyslexic Computer Users, *British Journal of Educational Technology*, Vol. 34, No. 3, pp.341-355.

James, C. L., & Reischel, K. M., (2001), Text Input for Mobile Devices: Comparing Model Prediction to Actual Performance, *Proceedings of CHI2001*, ACM, pp. 365-371, New York: ACM.

Jongmans, M. J., Linthorst-Bakker, E., Westenberg, Y., & Smits-Engelsman, B. C. M (2003), Use of a Task-oriented Self-Instruction Method to Support Children in Primary School with Poor Handwriting Quality and Speed, *Human Movement Science,* Vol. 22, pp. 549-566.

Keates, A., (2000), *Dyslexia and Information and Communications Technology: a Guide for Teachers and Parents*. David Fulton Publishers.

MacKenzie, I. S., (2002), Introduction to This Special Issue on Text Entry For Mobile Computing, *Human-Computer Interaction,* Vol.17*,* pp.141-145

MacKenzie, I. S., & Chang, L., (1999), A Performance Comparison of Two Handwriting Recognisers, *Interaction with Computers*, Vol. 11, pp.283-297.

MacKenzie, I. S., & Soukoreff, R. W., (2003), Phrase Sets for Evaluating Text Entry Rate Development, *Extended Abstracts of the ACM Conference on Human Factors in*

Computing Systems – CHI 2003, pp.754-755. New York: ACM.

MacKenzie, I. S., & Soukoreff, R. W., (2002), Text Entry for Mobile Computing: Models and Methods, Theory and Practice, *Human-Computer Interaction*. Lawrence Erlbaum Associates. Vol. 17, pp. 147-198.

McHale, K., & Cermak, S. A., (1992), Fine Motor Activities in Elementary School: Preliminary Findings and Provisional Implications for Children with Fine Motor Problems, *American Journal of Occupational Therapy,* Vol. 46, pp. 898-903.

Miller, N. (1986), *Dyspraxia and its Management*, London: Croom Helm.

Miozzo, M., & Bastiani, P. D., (2002), The Organisation of Letter-Forming Representations in Written Spellings: Evidence from Acquired Dysgraphia, *Brain and Language*, Vol.80, pp.366-392.

Newell, A. F., Booth, L., and Beatie, W., (1991), Predictive Text Entry with PAL and Children with Learning Difficulties, *British Journal of Educational Technology*, Vol. 22, No. 1, pp.23-40.

Ott., P., (1997), *How to Detect and Manage Dyslexia, A Reference and Resource Manual,* Heinemann Educational Publishers

Read, J. C., MacFarlane, S. J. and Casey, C. (2001) Measuring the Usability of Text Input Methods for Children, *presented at HCI2001*, Lille, France, Vol. 1 Springer Verlag, pp. 559 – 572

Silfverberg, M., MacKenzie, I. S. & Korhonen, P., (2000), Predicting Text Entry Speed On Mobile Phones, *Proceeding of CHI2000,* ACM, pp. 9-6 New York; ACM

Soukoreff, R. W., & MacKenzie, I. S., (2004), Recent Development in Text Entry Error Rate Measurements, *Extended Abstract of the ACM Conference on Human Factors in Computing Systems*, pp. 1425-1428, New York, ACM.

Soukoreff, R. W., & MacKenzie, I. S. (2003). Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. *Proceedings of the ACM Conference on Human Factors in Computing Systems* – CHI 2003, pp. 113-120

Waites L., (1968), Report on research group on developmental Dyslexia and world illiteracy, *Dyslexia International World Federation of Neurology, Bulletin of the Orton Society* Vol. 18, pp. 21-22