Deceit, Distortion and Decision – Choosing Phrase Sets for Text Entry Research

Janet C Read

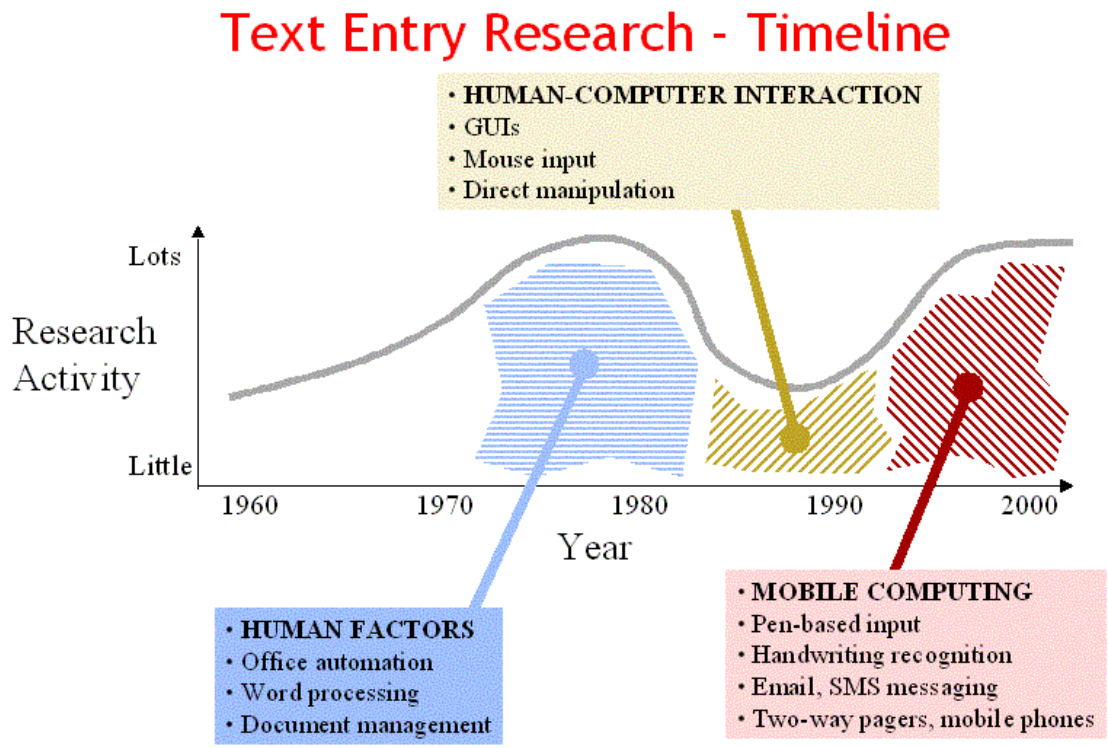
Abstract

This paper describes the design and early evaluation of a method for validating phrase sets for text input experiments. The paper begins with a historical overview of text input research and then widens the discussion to consider the effects on the research results of different users and different technologies. A framework is proposed for future verification of phrase sets and a small empirical study is described that provides evidence that the framework has potential.

1 Text Input Research

We spend a significant amount of time carrying out text input activities. This time is generally made up of input time and correction time. Our time and effort would be reduced if it were possible to have more efficient text input methods that were quicker to use and provided improved accuracy and error repair.

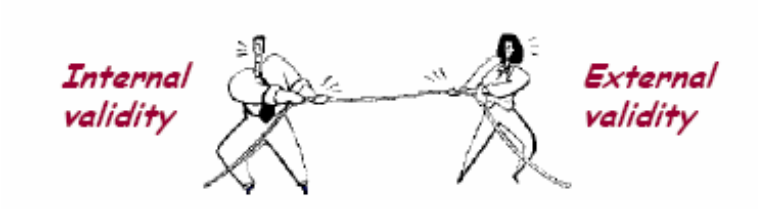
The study of text input has been around for some time.



It is usual to make judgements of text input by using text phrases that users copy. In a traditional experiment, the user is presented with five or more phrase and the researcher measures the time taken to copy the words, the errors made, and in some instances, the time taken for repair. Other measures that are sometimes taken include measures of uncorrected errors and counts of keystrokes.

Given that the results from these experiments are then used as evidence of the usability of text input methods and interfaces, it is important to consider what (if any) affect the choice of text phrases will have on the experimental results. Much of the work in this area has been done by (MacKenzie and Soukoreff 2003) who have produced the key papers in this area, proposing that researchers need to take phrase selection seriously and providing free web based tools that allow text phrases to be compared with one another and compared with language models. These comparisons allow the researcher to claim that their text phrases match the language models of the users and also allow researchers to claim that within a set of phrases there is consistency. (Soukoreff and MacKenzie 2003) have produced and published a list of text entry phrases in the hope that researchers might use this list and thereby allow for comparisons across research studies that are carried out in many places. Some of these phrases have recently been used by (Read 2005) in a study of the usability of digital ink technologies with children.

There is a great trade off in text input research between internal and external validity. To maintain internal validity it is necessary to apply many controls to the experimental design, this often requires participants to copy phrases that are chosen by the researchers , Copying text is not a natural activity for most of us, and so by using copied text for text input experiments, external validity is compromised as the work does not match the real world particularly well.



In most of the work on text input, researchers tend to use copied text phrase as those produced by (MacKenzie and Soukoreff 2003).

2 Constructing Text Input Phrases

When a researcher is about to embark on a text input experiment there is a series of stages through which his design process takes him. The purpose of a text input activity is important. It is essential to understand that if a user is carrying out the text input while moving, or in low light it will be different from text input at a well lit desk. Also, the purpose of the text input activity is important. Text messaging and text input in a chat area is different from text input at a word processor.

- What is the language and context of the study?
- Who are the users?
- What is the technology?

Each of these has a bearing on the text phrases that are being used; they are discussed here.

2.1 Language and Context

This is a well understood area of text input research.

The language that is being used in the study is not just about the name of the language, e.g. English, but also about the style of the language, that is, is it written language or spoken language. There are a number of approaches to language modelling; one is to compare the language to a frequency model, matching the frequencies of letters and words within the text phrase to the frequency of letters and words in the language. This method uses a corpus of language, for example the (BNC 1995). A common approach is to check to see if all the letters of the alphabet exist in the phrase set; this is less useful than a language model and reduces external validity.

2.2 Users

'*All people are equal, but some are more equal than others*' (Orwell 1945). It is not possible to control the selection of users at an experiment but the selection of text phrases needs to reflect the user population. (MacKenzie and Soukoreff 2002) suggests that text entry research is more valid if the users are able to remember the text phrase and then enter it without having to refer back and forth to the written phrase. This reduces the Focus of Attention by one and provides a closer match to real world text input activities. When the intended users are children, or older people, or people with some language difficulty (this may include conditions such as dyslexia as well as having the presented language as a second language rather than a first).

2.3 Technology

When a text input experiment is being carried out on a single technology, perhaps to see how users of different ages or how users in different conditions perform, it is acceptable to disregard the technology in the design of the text phrases. However, when phrases are being used to compare performance across different technologies, it is probable that in some technologies, some phrase may be more difficult to replicate than others. An example of this is the production of the word 'cab' and the production of the word 'sad' on a mobile phone using multitap.

3 Text Phrase Selection

Some of the text phrases that have been proposed by (MacKenzie and Soukoreff 2003) include the following:

My watch fell in the water
Time to go shopping
You must be getting old
The world is a stage
Do not say anything
Are you talking to me
You are very smart
All work and no play
Did you have a good time

3.1 Assessing the risk in text phrases

On the assumption that not all text phrases are equal, an study was proposed to determine the effect of risk factors on phrase selection. In the following table, just by identifying two 'risky' letters for handwriting (I, and n) and by noting those sequences of letters that appear on the same key in a multi-tap texting environment, it can be seen that different phrase have different risks in different instances. The final column shows the trigger words

	H		M		TW	
My watch fell in the water	in	2	Fe ll	2	My In the	3
Time to go shopping	I in	3	pp	1	To go	2
You must be getting old	in	2	tt	1	You Be getting	3
The world is a stage	i	1		0	The A	2
Do not say anything	N in	3	No hi	2	Do Not anything	3
Are you talking to me	in	2	lk	1	You To me	3

It is hypothesised that when using a keyboard, multiple letters, e.g. ll and adjacent letters, e.g. gh might be more prone to errors that more spaced out letters. When writing with a pen, there are some letters that are known to be more erroneous in recognition, these include I, n, v as well as the pairs cl and ai.

3.2 Devising an Experiment to determine the effect of risk in Multitap

This experiment is described by way of an example to show how further studies into this might progress. Multitap is used in this example, as it is quite well understood. A parallel study would be needed to look at handwriting or keyboard use and make some comparisons. This is described in the conclusion.

3.2.1 Determine the risk to be investigated.

In this instance, a single risk was identified, that being the consecutive use of a single physical key. This is known to be more difficult (MacKenzie and Soukoreff 2002) than texting using distinct keys, mainly because of the need to wait between key presses.

3.2.2 Construct text phrases with variable risks

From the phrase list proposed by (MacKenzie and Soukoreff 2003), the following phrases were selected;

	Risk	Risk Factors
I will meet you at noon	High	Ll, ee, no, oo, on
Take a coffee break	Medium	Ff, fe, ee
The world is a	Low	

stage		

3.2.3 Carry out a user test

The user test would require at least 12 participants to input these phrases, arranged in different orders to avoid learning effects. The time taken, and the number of errors in each would be measured.

3.2.4 Extending to cross technologies

To see how (if at all) these risk factors affected cross technology studies, it may be necessary to work with a set of nine phrases as follows, or it may be sufficient to use just six phrases by eliminating the medium cases.

	Technology 1	Technology 2
Phrase 1	High	High
Phrase 2	High	Medium
Phrase 3	High	Low
Phrase 4	Medium	High
Phrase 5	Medium	Medium
Phrase 6	Medium	Low
Phrase 7	Low	High
Phrase 8	Low	Medium
Phrase 9	Low	Low

Participants could be either one cohort split into two equal groups, one group doing Technology 1 then Technology 2, the other group doing Technology 2 then Technology 1, or they could be different groups in two places (less valid) – each doing a different technology.

Within each group – we would ideally use more than one phrase at each risk level, and each would do all six, or nine phrase possibilities.

4 Conclusion

This paper has proposed a research study that will be completed in the summer. It has outlined some of the problems with the use of phrases for text input.

5 References

BNC (1995). British National Corpus, Oxford University Computing Services. **2004**.

MacKenzie, I. S. and R. W. Soukoreff (2002). "Text Entry for Mobile Computing: Models and Methods, Theory and Practice." Human-Computer Interaction **17**(2): 147 - 198.

MacKenzie, I. S. and R. W. Soukoreff (2003). Phrase Sets for Evaluating Text Entry Techniques. CHI 2003, Ft. Lauderdale, FL, ACM Press.

Orwell, G. (1945). Animal Farm, Penguin books.

Read, J. C. (2005). The Usability of Digital Ink Technologies for Children and Teenagers. HCI2005, Edinburgh, Scotland, Springer.

Soukoreff, R. W. and I. S. MacKenzie (2003). Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. CHI2003, Ft. Lauderdale, ACM Press.