

An eye-tracking study on the importance of consistent terminology

Jacob Gavin*, Jimmy Hammarbäck*, Madeleine Hammarbäck*,
Benjamin Helmersson*, Martina Nyberg*, Cassandra Svensson*,
Jody Foo[†], Felix Koch[†]

Linköping University
Linköping, Sweden

*{jacga987, jimer336, madfr972}@student.liu.se

*{benhe882, marny531, cassv004}@student.liu.se

[†]{jody.foo, felix.koch}@liu.se

Abstract

Using inconsistent terminology, e.g. having different terms in documentation and e.g. labels and menu items in a user interface is believed to be confusing to users. However, few empirical studies exist on this particular topic. In this paper we show how users' interaction with an interface is affected by inconsistent terminology. An experimental eye-tracking study with 30 participants was conducted where the participants were shown a user interface and a task description. The terminology in the interface and task description was manipulated to be either consistent or inconsistent. The results show that terminological inconsistencies led to a significantly higher number of visual fixations, more time needed to perform the task, and more returns to the task description. The conclusion is that inconsistent use of terms create unnecessary cognitive workload for the user that can be avoided by ensuring terminological consistency within a system.

1. Background

As of today, most computer software uses written instructions to describe how to interact with the interface. It is a common belief that a consistent use of terms should be applied within the given domain in both interface design (Nielsen, 1994; Johnson, 2010) and translations (Merkel, 1996). Inconsistent use of terms between task description and interface can be said to occur when a term from the task description cannot be found in the interface. This inconsistency may have consequences for how a user interacts with the interface. Studies on the impact of terminological inconsistency are hard to find. Studies have been conducted on inconsistency, but the focus in these studies has not primarily been on terminology, but rather on the visual elements in the interface i.e. (Ozok and Salvendy, 2000). Cohen (2004) also arrive at the conclusion that consistent terminology should be used in voice interfaces.

2. Problem formulation

The number of errors made and the time needed for executing a task are common measures of usability. The idea of measuring the user's eye movement is to examine relatively easy tasks where users make few errors. In this study the number of fixations is used as a measure of visual focus, where more fixations means that the user needs to search more extensively to find the relevant information and can be assumed to lead to a higher cognitive load. Time for completing a task is also measured in this study, which is a common measure of usability. Additionally, this study also measures the number of times the user returns to the task description during the visual search. More returns to an area indicate semantically informative areas and can be seen as a need to reconfirm some information before final judgement.

The aim of this study is to examine the significance of consistent terminology between task descriptions and inter-

faces, using eye-tracking as a method. It is predicted that inconsistent terminology between task descriptions and interfaces will lead to (1) a higher number of eye fixations in the area of the task description and of the interface, (2) a longer amount of time spent on executing the task specified by the description, and (3) a higher number of returns to the area of the task description, compared to consistent terminology.

3. Method

To examine the significance of consistent terminology, an experiment was designed where participants performed tasks in interfaces by following task descriptions while their eye movement data was collected. The task description and the interface were displayed simultaneously.¹ The independent variable was the relation between the term in the interface and the term in the task description. This relation could be either *consistent* or *inconsistent*. The dependent variables were the number of fixations made in the area containing the task description and the interface before a decision was made, total amount of time spent on executing the task specified by the description, and the number of returns to the task description. Participants ($N=30$, $M=23.3$ years, $SD=2.4$ years) consisted of 15 female and 15 male students having Swedish as their first language. The participants performed 32 short visual search tasks. A repeated measures design was used, where half of the tasks used consistent terminology and the other half used inconsistent terminology. The interfaces and task descriptions were counter-balanced across consistent and inconsistent tasks. Half of the participants performed the tasks in reverse order. Thus, any impact of the interfaces or terms selected as well as learning effects was controlled for.

With support from SICS East ICT AB

¹Both task description and interface language was Swedish

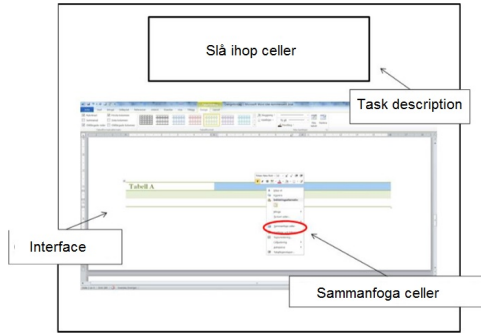


Figure 1: Illustration of a task with inconsistent terms.

4. Stimulus material

In carrying out the study, participants were shown a total of 32 screenshots from 16 different interfaces with different task descriptions located above (see Fig. 1). Eye movement, mouse clicks and task duration were recorded. Eight pairs of terms (16 unique terms) were distributed between task descriptions and interfaces. Each pair of terms P consisted of 2 terms, T_1 and T_2 , belonging to the same concept. The use of terms in the material varied between consistent and inconsistent. Consistent use of terms meant that one of the terms in a pair P was used in both the task description and the interface. Inconsistent use of terms meant that one of the terms in the pair P was used in the task description, and the other term in the pair P was used in the interface.

5. Analysis

A paired-sample t-test was used to examine the difference between consistent and inconsistent terminology with regard to (1) the number of fixations until a decision was made, (2) the time required to complete the task, and (3) the number of returns to the task description in both consistent and inconsistent conditions.

6. Results

The result of the data analysis (see Table 1) showed that the tasks with inconsistent use of terms ($M=34.9$, $SD=9.6$) yielded a significantly higher number of fixations than tasks with consistent use of terms ($M=23.8$, $SD=8.3$, $t(29)=6.35$, $d=2.36$, $p<.001$). Furthermore, it was shown that tasks with inconsistent use of terms ($M=15.7$, $SD=4.1$) required significantly more time per task compared to tasks with consistent use of terms ($M=10.8$, $SD=2.8$, $t(28)=6.90$, $d=2.60$, $p<.001$). The result from the analysis also indicate a significantly higher number of returns to the task description when the use of terms was inconsistent ($M=1.13$, $SD=.47$) compared to consistent use of terms ($M=.84$, $SD=.53$, $t(27)=-3.19$, $d=-1.23$, $p=.004$).

		M	SD	t	Cohen's d
Fixations	Inconsistent	34.9	9.6	$t(29) = 6.35^{***}$	2.36
	Consistent	23.8	8.3		
Required time	Inconsistent	15.7	4.1	$t(28) = 6.90^{***}$	2.60
	Consistent	10.8	2.8		
Returns	Inconsistent	1.13	0.45	$t(27) = 3.19^{**}$	1.23
	Consistent	0.84	0.53		

Table 1: Paired-sample t-test results. ** $p<.01$, *** $p<.001$

7. Discussion and conclusions

The hypothesis of differences between inconsistent and consistent use of terms in the task description and interface was tested and indicated a statistical significance for both time per task and number of fixations. The results also showed that there was a higher number of returns to the task description when the use of terms was inconsistent. The difference is rather small, which could indicate that the participants kept the task description in their working-memory throughout the task, regardless of the condition.

When using the same term in the task description as in the interface both the number of fixations and task duration is affected. Fewer fixations are needed before a decision is made and a shorter amount of time that is spent solving the task. While reading a task description, a mental model is created of the task ahead. It is likely that the mental model partly consists of key terms used in the task description and that the user uses these to find the solution in the interface. Since the study shows a higher number of fixations and an increased amount of time when having inconsistent terminology between the task description and the interface compared with a consistent terminology it is plausible that the key terms in the mental model are more like concrete representations than abstract concepts. Hence, a reasonable interpretation could be that the terms are stored in a part of the working memory, the phonological loop, as auditory memories. This idea is underpinned by earlier studies (Baddeley, 2006).

The participants used an average of almost 33% more time to solve tasks involving inconsistent terminology, even though the tasks were relatively easy. How this increase in time is reflected in more complex scenarios needs to be studied further. However, the experimental context can be compared to an inexperienced user interacting with an interface for the first time or after a long period of time since previously using it. Terminological and linguistic consistency in systems can be assured using language technology such as authoring support and properly maintained translation memories, which would reduce the cognitive load of end users when reading these texts.

References

- A. D. Baddeley. 2006. Working memory: an overview. In Pickering S., editor, *Working Memory and Education*, pages 1–31. Academic Press.
- M. H. Cohen, J. P. Giangola, and J. Balogh. 2004. *Voice User Interface Design*. Addison Wesley, Redwood City.
- J. Johnson. 2010. *Designing with the Mind in Mind*. Morgan Kaufmann.
- M. Merkel. 1996. Consistency and variation in technical translation: A study of translators' attitudes. In D. Kenny L. Bowker, M. Cronin and J. Pearson (ed.), editors, *Unity and Diversity? Current Trends in Translation Studies*, pages 137–149. St. Jerome Publishing.
- J. Nielsen. 1994. Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM.
- A. A. Ozok and G. Salvendy. 2000. Measuring consistency of web page design and its effects on performance and satisfaction. *Ergonomics*, 43(4):443–460.