Slips of the Keyboard

Kurolt, Silvia

University of Zagreb
Faculty of Humanities and Social Sciences
Department of English
English Language Section

Master's thesis

Slips of the Keyboard:
Identification and Classification of Typing Errors made by Native Speakers of English

Candidate: Silvia Kurolt
Advisor: Prof. Irena Zovko Dinković
Academic year: 2022/2023

Zagreb, February 2023

Table of contents

1. INTRODUCTION

This thesis is being written on a personal computer. All sources used are currently open on the second monitor of that same personal computer. On the table there is also a notebook with a pen beside it. But in it, only 5 or so lines are written down. The rest of the messy notes, all of the disjointed thoughts that have yet to grow into fully formulated sentences and paragraphs, are roughly typed out in an entirely different word processing program. The entire process is done with a keyboard. This is not unusual, nor is it rare. Today, many people rely on electronical devices such as personal computers and mobile devices to complete their everyday tasks. Everything from socializing, keeping up with the news and entertainment, to note-keeping and habit-tracking falls within the capabilities of these smart devices. And to top it all off, "the average person spends a total of 6 hours and 57 minutes looking at a screen each day" (Moody, 2022), 3 hours and 14 minutes of which are spent looking at a computer screen (Moody, 2022). And while those 3 hours and 14 minutes of computer screen time do not directly translate into 3 hours and 14 minutes of typing, the real number likely isn't too far.

So, what does this have to do with the so-called *slips of the keyboard*? Well, a lot. Typing errors or slips of the keyboard have been around for as long as there have been keyboards and despite many people in the modern day having adjusted to the daily use of various keyboard types – be they connected to our personal computers or displayed on the touch screens of our mobile devices – typing errors are showing no signs of going away. It could even be argued that they have become even more prevalent and common as a result of the growing role of technology in our everyday lives, workspaces, schoolwork, etc. This assumption is supported by the existence of numerous spelling and grammar checkers, and especially the little wavy red line that appears under incorrectly spelled words in seemingly every application on every device. And yet, when searching for studies examining different kinds of linguistic slips, it is clear that it is actually the slips of the tongue that capture the interest of most linguists. A wide selection of studies on typing errors is available, but a majority of these are from a non-linguistic perspective and often for the purpose of the development and improvement of tools such as the aforementioned spelling checkers and grammar checkers. It is certainly more difficult to find studies examining the actual process of typing and especially the process of erring while typing from a (psycho)linguistic perspective. This is in line with what Harley (2001) observed about research on language production in general – that "[t]here has been less research on language production than on language comprehension" (p.

374). This observation is partially explained by the fact that "[u]nlike speech, reading and writing are a relatively recent development" (p. 193) – though concrete evidence is limited, the one widely accepted theory suggests that the human ability for speech, and subsequently language, emerged approximately 200,000-300,000 years ago (Gutman-Wei, 2019), whereas "[t]he first writing system […] appeared just before 3000 BC" (Harley, 2001, p. 193). Compared to 200,000-300,000 years of speech and some 5000 years of writing, typing has been around for only about 300 years old as the first writing device of this kind was patented in 1714 by Henry Mill (Computer Hope, 2021). It is important to note that while the history of typing does officially begin in 1714, the first successful modern typewriter was actually released in 1896 and sold five million pieces by 1939 (Computer Hope, 2021). But it wouldn't be until the widespread of personal computers and the Internet in the late 1990s ("Diffusion of Personal Computing Devices, 1977-2020," 2022) that typing would become a regular, everyday activity and soon a widely taught (and expected) skill. So, to reiterate, while it is not at all surprising that speech has gotten significantly more attention from linguists, especially considering the 300,000-year head start it had on typing, it is at the very least unusual that one of the most frequently used language output systems today hasn't garnered more interest – which is exactly why this topic piqued my interest, especially as someone whose average daily screen and/or keyboard time well surpasses the above quoted 3-hour and 14-minute estimate.

Therefore, in this thesis I present an analysis of a typing error corpus collected from samples provided by 25 participants whose first language is English. The first part of the thesis provides the theoretical framework for the study and an overview of previous studies done on this topic. The second part of the thesis contains a detailed description of the methods used to collect and prepare the data used in the study. In the third and final part of the thesis results are presented, analyzed, and compared to previous research.

2. RELATED RESEARCH

The term *language production* and its counterpart, *language comprehension* were previously briefly mentioned without much elaboration. However, in order to better understand how typing fits within the generally recognized language skill framework, the two terms need further explaining. The framework consists of four main language skills or behaviors: speaking, listening, writing, and reading (Harley, 2001, p. 422). Each of the skills can be categorized as visual or verbal, and input or output as shown in Table 1 (p. 421). Based on this categorization, *language production* refers to the two output skills, whereas *language comprehension* refers to the input skills.

Table 1

Four Language Skills (adapted from Harley (2001: 421))

	Input	**Output**
Visual	reading	writing
Verbal	listening	speaking

This framework usually does not mention typing as a separate language skill, but typing is sometimes discussed in relation to the writing skill based on the similarities in both the psychological and physical processes of typing and writing (Berg, 2002; Pinet, Ziegler, and Alario, 2016). Others, such as MacKay (1993), consider typing a separate output modality (p. 66). Regardless of strict classification, typing is commonly observed separately and, just like speech and writing, considered a complex process that occurs in stages. For each of the three processes – speech, writing, and typing – "noise in any of the stages or connecting channels […] can distort the original message" (Fromkin, 1973, p. 182) resulting in an error or *slip*. These errors can then be categorized into speech errors (slips of the tongue), writing errors (slips of the pen), and typing errors (slips of the keyboard). The following is an overview of some of the studies dealing with the slips of the key/keyboard.

MacNeilage (1964) stands as one of the earliest attempts at collecting, classifying, and analyzing a set of typing errors with the aim of understanding the underlying mechanisms involved in the process of typing. The observed dataset consisted of "623 typing errors produced by 5 [participants]" (p. 144) using the typewriter. Not having the luxury of a keylogger[1] at his disposal, MacNeilage instead opted for a different way of collecting what would otherwise have been a corrected error corpus – no error correction was allowed, instead, when the participants made an error "[t]hey were instructed to stop typing […] and to begin again at the beginning of the word in which the error had been made" (p. 145). The error classification he proposes in this paper is made up of five major types with multiple subtypes each. A detailed breakdown of the classification can be seen in Table 2, whereas Table 3 provides examples for each error (sub)type.

Table 2

Typing Error Classification with Detailed Descriptions (adapted from MacNeilage (1964: 146-147))

Main error types	Description	Error subtypes	Description
Spatial errors	Errors resulting from typing a letter immediately adjacent on the keyboard to the one required by the copy	**Horizontal errors**	Typing a letter immediately above or below the correct letter in the same row of the keyboard
		Vertical errors	Typing a letter immediately above or below the correct letter in the same column of the keyboard
		Diagonal errors	Tying a letter in a row and a column adjacent to that of the correct letter
Temporal errors	Errors in the order in which the required letters were typed	**Reversal errors**	Reversing the order of two letters next to each other in the correct sequence
		Omission errors	One letter in a sequence is left out

[1] Today, the term *keylogger* or *keystroke recorder* is mentioned in the context of information security and data theft. It is most commonly defined as "a piece of software that records the signals sent from a keyboard to a computer usually for the purpose of gaining information about the user without the user's knowledge" (Merriam-Webster, n.d.), but this definition is lacking in the context of typing research and even more so when discussing studies preceding the personal computer boom of the 1990s. Therefore, a more relevant definition would define a keylogger as any piece of software or hardware (the likes of which was used in Tannenbaum, Williams & Wood (1967)) that is used to record all strokes made on a keyboard. A keylogger can also optionally record time intervals between individual keystrokes.

		Equivocal errors	The letter one stroke ahead of the one required is typed, after which the participant stops (becoming aware of the error)
		Anticipation errors	A letter is typed which is required more than one stroke ahead of the place where it is mistakenly typed
Miscellaneous errors	A number of other specific error types	**Interpolation**	A letter apparently quite unrelated to the correct sequence is inserted in it
		Phonemic errors	A letter pronounced similarly to that of the correct letter replaces the correct letter
		Type errors	One letter of a word is changed, making it into a word similar to the correct one but meaningless in context
		Contralateral errors	A stroke is typed using the same row and corresponding finger to the correct one, but with the other hand
		Dynamics errors	The letter adjacent in the sequence to a letter which should have been typed twice, is typed twice instead
Multiple classification errors	Errors which can be placed in more than one category	**-**	-
Unclassifiable errors	Errors which could not be placed into any of the above categories	**-**	-

Table 3

Typing Error Classification with Examples (adapted from MacNeilage (1964: 146-147))

Main error types	Error subtypes	Example
Spatial errors	Horizontal errors	“e” instead of “r”; “d” instead of “f”
	Vertical errors	“f” instead of “r”; “e” instead of “d”
	Diagonal errors	“d” instead of “r”

Temporal errors	Reversal errors Omission errors Equivocal errors Anticipation errors	"ht" instead of "th" "lenth" instead of "length" "stiml-" instead of "stimulus" "ext-" instead of "expected"
Miscellaneous errors	Interpolation Phonemic errors Type errors Contralateral errors Dynamics errors	"forme*n*d" instead of "formed" "ma*c*" instead of "makes" "tha*t*" instead of "than" "tr*e*als" instead of "trials" "er*o*ors" instead of "errors"
Multiple classification errors	-	"re*s*pression" instead of "repression" (the first "s" could be an interpolation or an anticipation error
Unclassifiable errors	-	"condi*d*ioning" instead of "conditioning"

The results of his analysis showed that "70% of all errors was classifiable into particular single categories within the main classes of spatial, temporal and miscellaneous" (p. 147), "10% was not classifiable into any category" (p. 147), and "20% was classifiable into more than one category" (p. 147). He observed that spatial errors "occurred significantly more often on the least frequent letters in the language" (p. 155) and were in general "affected by keyboard difficulty variables" (p. 156). Additionally, the same factors that affected the spatial errors – "[h]andedness, keyboard structure and letter frequency" (p. 156) – did not significantly affect the temporal ones. Temporal errors were instead affected by digram frequency and word length (p. 156). However, both spatial and temporal errors were affected by word boundary. MacNeilage concludes that spatial and temporal errors mirror two different mechanism that are at work during typing, the "executive" one and the "programming" one, respectively (p. 157). The "executive" mechanism is characterized by "its sensitivity to the physical difficulty of the behavior" (p. 157), whereas the "programming" mechanism shows "sensitivity to certain variable of language structure" (p. 157). Finally, he proposes a three-stage production mechanism for typing which consists of "(1) determining tendency (after Lashley[, 1951]), (2) programming mechanism and (3) executive mechanism" (p. 158).

In 1985, MacNeilage expands on his previous work with the aim of comparing "errors in speech and typing" (p. 193) relying on the error corpus and classification presented in his 1964 paper. The data is compared to speech errors, which MacNeilage classified into five types: (1)

substitution, (2) addition, (3) omission, (4) shift, and (5) exchange (p. 194). Harley (2001) however provides a more detailed speech error classification consisting of 11 error types, as shown in Table 4 (p. 377). Of these five, all types but shift errors have analogs in typing error types (p. 194). Additionally, "substitution errors [were] not considered […] [despite occurring in both speech and typing], because typed substitution errors seem to primarily involve spatial contingencies (typing a letter spatially adjacent to the correct one) rather than temporal factors" (p. 194). As for the typing error types, of the original five main types and 12 subtypes outlined in MacNeilage (1964), only one main type (split into four subtypes) and one additional subtype are used for the comparison. The exact error (sub)types used for the comparison and their speech analogs are shown in Table 4.

Table 4

Comparison of Error Types in Speech and Typing (adapted from MacNeilage (1985: 194))

Speech error type	Typing error type
Addition	Interpolation[2]
Omission	Omission[3]
Exchange	Reversal
-	Equivocal
-	Anticipation

What MacNeilage (1985) observes is that speech errors tend to "conform to the phonotactic rules of the language" (p. 195), especially within the "internal structure of syllables" (p. 195), something that cannot be said for the typing errors he observed and analyzed in his research (p. 195). Special attention is paid to the frequency of consonant vs. vowel addition and omission in both speech and typing. MacNeilage observes that in speech "neither consonant additions nor omissions change syllable number" (p. 197) and therefore don't have "severe repercussions on syllable structure" (p. 197). The opposite is true when it comes to vowel addition and omissions in speech – they "change syllable number, and vowel omissions [specifically] may result in impermissible syllable structures" (p. 197). As might be expected with what was already noted about typing errors in general, "there appears to be no constraint against the addition or omission

[2] In MacNeilage (1964), this is a subtype of the miscellaneous error type (p. 146)
[3] In MacNeilage (1964), this is a subtype of the temporal error type (p. 146), as well as the reversal, equivocal and anticipation error type

of vowel letters in typing" (p. 197). Overall, MacNeilage's comparison makes it very clear that despite the fact that both speech and typing deal with the same language, the two output systems will often produce significantly different sets of errors, pointing to the differences in the very process of production for each system, which is also evident from a more elaborate classification of speech errors by Harley (2001: 377), who distinguishes 11 different types of speech errors[4].

Another study which intended to build upon MacNeilage's original 1964 work was a paper by Shaffer (1975). It too was inspired by preexisting studies of speech errors, and it aimed to classify and explain typing errors in a fairly similar way. The study involved only one participant who was, similarly to the 1964 study by MacNeilage, "instructed not to correct her errors" (p. 422). The results of the experiment are interpreted within the context of a two-stage typing process proposed by Shaffer: "[i]t is assumed that the first stage translates its string simultaneously and deposits the response codes in a buffer memory, while the second stage converts response codes iteratively into finger movements" (p. 420). Shaffer divides the resulting corpus of around 440 errors into four categories titled "Omit, Alter, Add, and Transpose" (p. 422). In the error analysis, a few observations were of special note. Firstly, Add and Transpose errors occurred almost exclusively towards the end of a word. Secondly, although the importance of a syllable is severely reduced in typing compared to its importance in speech, some errors exhibited "an exchange of letters between homologous syllabic positions" (p. 423), as well as syllable boundary crossing. This points to "a conclusion that syllabic structure affects typing output" (p. 423) but should be noted that there were no cases of "migration of a whole syllable" (p. 423), which tends to occur in speech errors.

A study by van Nes (1976) examined a collection of 293 errors made by 25 typists, which were further classified into seven error categories. The basis for classification was "a three-fold

---

[4] The 11 error types were the following: **feature perseveration** (e.g. "Turn the knop" for "Turn the knob"), **phoneme anticipation** (e.g. "The mirst of May" for "The first of May"), **phoneme perseveration** (e.g. "God rest re merry" for "God rest ye merry"), **phoneme exchange** (e.g. "Do you reel feally bad?" for "Do you feel really bad?"), **affix deletion** (e.g. "chimney catch fire fire" for "chimney catches fires fires"), **phoneme deletion** (e.g. "Backgound lighting" for "background lighting"), **word blend** (e.g. "The chung of today" for "the children+young today"), **word exchange** (e.g. "whose mind came to name" for "whose name came to mind"), **morpheme exchange** (e.g. "I randomed some samply" for "I sampled some randomly"), **word substitution** (e.g. "Get me a fork" for "Get me a spoon"), **phrase blend** (e.g. "Miss you a very much" for "Miss you very much" and "Miss you a great deal") (p. 377)

division of processes that may be assumed to occur in keying" (p. 167) and which consist of (1) "identifying, reading and storing the information to be keyed from a source document" (p. 167), (2) "removing the information from a short-term store, and converting it in a series of movement commands to the appropriate fingers" (p. 167), and lastly (3) "actually depressing the target keys in the programmed order" (p. 167). Errors were then categorized depending on the stage at which an error occurred, with subcategories for each stage. When discussing error classification, van Nes very interestingly pointed out that there were usually "only one or two plausible causes for an error" (p. 168) and that true error ambiguity was rare.

A study by Grudin (1983) compared errors made by novice and skilled typists. The dataset used was collected with the help of six professional typists and eight novice ones. The analysis was mostly concerned with substitution errors, but all errors collected were classified into the insertion, omission, substitution, transposition, and 'other' categories. The analysis also touched upon the topic of "activation error" (p. 132) and the effects of surrounding within-word context on the error occurrence.

MacKay (1993) attempted what Berg would also attempt some years later – to compare and analyze typing and writing errors – though this study involved a three-way comparison with speech as well. The comparison was based on previous research, some of which has already been discussed in this chapter. A five-category classification was used: substitution errors, intrusion errors, omission errors, transposition errors, and other errors. The conclusion of the study was in line with previous observations by MacNeilage (1985) – speech errors differ significantly from both those in typing and in writing, primarily because speech itself differs so vastly from both typing and writing. Differences between typing and writing were also noted.

A study by Berg (2002) dealt exclusively with uncorrected typing errors. His corpus consisted of "500 typographical errors excerpted from scholarly works published in English" (p. 187). In discussing the similarities and differences between speech, writing, and typing errors, Berg pointed out that "typing appears to be more similar to writing than to speaking" (p. 186) as both output systems "make use of the hand, are acquired relatively late or not at all, require some formal instruction to be mastered, are largely based on the prior acquisition of speaking, create a visual-spatial code, and serve almost identical functions" (p. 186). The aim of Berg's study, then, was to compare typing and writing errors in order to see if the aforementioned similarities were enough to claim that the mental processes behind both typing and writing were the same. Berg offers an almost painfully thorough classification of errors. In an attempt to accurately present and explain this classification system, we provide a list of terms and their explanations in Table 5, whereas precise relationships between different monopositional error types and their descriptions are shown in Tables 6, 7, and 8. Table 9 covers the bipositional error classification and Table 10 expands on the classification of adjacent switches in bipositional errors. Examples given by Berg for some error types are also provided.

Table 5

List of Terms Used in Berg's (2002) Error Classification (adapted from Berg (2002))

Term	Explanation
Monopositional error	Errors that involve a single grapheme.
Bipositional error	Errors that involve two graphemes.
Contextual error	An error is said to be contextually motivated when a source unit that is identical to the error unit can be found in its vicinity.
Non-contextual error	An error is considered non-contextual when a source unit that is identical to the error unit cannot be found in its vicinity.
Within-word error	Contextual error where the error unit is not separated from the source unit by word boundary.
Between-word error	Contextual error where the error unit is separated from the source unit by word boundary.
Substitution	Descriptive error category where the correct grapheme is replaced by another one. This category can be contextual or non-contextual.
Addition	Descriptive error category where an extra grapheme is added into an otherwise correct sequence. Can be contextual or non-contextual.
Omission (deletion)	Descriptive error category where one grapheme of the correct sequence is omitted. Can be contextual or non-contextual
Masking	A subtype of contextual omission errors. It refers to an error resulting from one element blotting out an identical unit in the vicinity.
Adjacent masking	A subdivision of masking errors where both elements are right next to each other.
Nonadjacent masking	A subdivision of masking errors where both elements are not next to each other.
Plain contextual omission	Error resulting from omission (deletion) of a unit due to structural pressure from a neighboring syllable or word.
Anticipation	A division according to the linear order of the error and source unit. In this case, the error unit precedes the source.
Perseveration	A division according to the linear order of the error and source unit. In this case, the error unit follows the source.
Adjacent switch	A bipositional error type in which two adjacent graphemes switch positions.
Reversal	A bipositional error type in which two nonadjacent graphemes switch positions. Also known as nonadjacent switch.

Shift	A bipositional error type where a grapheme skips surrounding graphemes to a position it is not supposed to be occupying.
Doubling	An error where a grapheme is doubled when it is not supposed to be.

Table 6

Contextual Within-Word Error Classification for Monopositional Errors (adapted from Berg (2002: 189-193))

					Example
Contextual within-word errors	Substitution	Anticipation			"re*b*ember" for "remember"
		Perseveration			"eviden*de*" for "evidence"
	Omission (deletion)	Masking	Adjacent[5]		"agram_atic" for "agrammatic"
			Nonadjacent	Anticipation	"inte_preted" for "interpreted"
				Perseveration	"con_luded" for "concluded"
		Plain contextual omission			-
	Addition				"in*it*uitive" for "intuitive"

Table 7

Contextual Between-Word Error Classification for Monopositional Errors (adapted from Berg (2002: 189-193))

				Example
Contextual between-word errors	Substitution	Anticipation		-
		Perseveration		"some *s*ombination" for "some combination"
	Omission		Anticipation	-

[5] Berg does not outrightly state whether or not adjacent contextual within-word masking errors can be divided by directionality into anticipatory and perseveratory subclasses. However, since it would be impossible to determine which of the identical units masked the other one, the assumption is that directionality does not matter for this particular error class.

	(deletion)	Masking[6]	Perseveration	-
		Plain contextual omission		"p_ant peas" for "plant peas"[7]
	Addition			-

Table 8

Non-contextual Error Classification for Monopositional Errors (adapted from Berg (2002: 189-190))

		Example
Non-contextual errors	Substitution	"*d*roups" for "groups"
	Omission (deletion)	"phonologi_al" for "phonological"
	Addition	-

Table 9

Bipositional Error Classification (adapted from Berg (2002: 194-197))

				Example
Bipositional errors	Adjacent switch	$C_1C_2 \rightarrow C_2C_1$		"a*ps*ect" for "aspect"
		$V_1V_2 \rightarrow V_2V_1$		-
		$CV \rightarrow VC$		"p*re*severate" for "perseverate"
		$VC \rightarrow CV$		-
		Grapheme doubling	Doubling after adjacent switch	"*eww*kends" for "weekends"
			Misordered adjacent grapheme doubling	"impo*sii*ble" for "impossible"
		Single grapheme		"ph*eno*me" for "phoneme"

[6] Berg does not specify whether contextual between-word masking can be adjacent and nonadjacent in the same way that contextual within-word masking can. However, it seems that all contextual between-word masking errors would be considered nonadjacent by default as no matter the position of the source and error unit, they would always be separated by the word boundary. Therefore, contextual between-word masking errors have no other subdivisions outside of those denoting directionality.

[7] Berg notes that this example "is a speech error [and] not a typographical one" (p. 191)

	Reversal (nonadjacent switch)	Grapheme doubling	Doubling after reversal	"co*nnum*ication" for "communication"
			Misordered nonadjacent grapheme doubling	"pro*ferr*ed" for "proffered"
	Shift	Anticipation		"pat*n*iets" for "patients"
		Perseveration		-

Table 10

Further Classification of Adjacent Switches in Bipositional Errors Based on Within-Word Position (adapted from Berg (2002: 195-196))

	Within-word position
Adjacent switches	Word-initial
	Postinitial
	Medial
	Pre-final
	Word-final

All aspects of the classification were used in a three-way comparison between speech, writing, and typing, leading Berg to conclude that "slips of the typewriter key are largely indistinguishable from slips of the pen [and] [b]y implication, the psychological processes underlying typing and writing are highly similar" (p. 204).

Among the more recent studies of typing errors, Baba and Suzuki (2012) carried out a study of typing errors in English and Japanese readers. In the study, a custom-built keylogger was used, which allowed for fine-tuning to fit the needs of the study and avoid issues that we describe in Chapter 3 of this thesis. The keylogger was used to collect a dataset of "corrected errors" (p. 373) which were then classified into "four descriptive types: Deletion, Insertion, Substitution and Transposition" (p. 375). The classified errors were then compared based on their frequency in both the corrected error and uncorrected error dataset. This way, they could precisely detect which factors played a role in a typist's successful or failed error detection – type of error, error positioning in the word, surrounding context, visual or phonological similarity, etc.

Last but not least, we refer to two notable studies which deal with typing from a psycholinguistic perspective but differ from previously mentioned research and this thesis, either in methodology or in the overall aim.

The first of these is a study by Tannenbaum, Williams and Wood (1967), dealing with hesitation phenomena in speech and typing. Of special interest is the method used to collect the typed part of the dataset – “[a] special typewriter, equipped to signal key striking, spacing, back-spacing, and carriage-return was used” (p. 205). And “[b]y linking the typewriter to a chronograph, a complete temporal record of all typing events […] was provided” (p. 205). Additionally, it was observed that whereas “[s]peech was generally characterized by relatively short bursts of output with frequent but brief interruptions” (p. 210), “[t]yping was more represented by fewer pauses of longer duration, each followed by relatively extended sequences of uninterrupted production” (p. 210). The study otherwise analyzes the aforementioned hesitation phenomena, which fall outside of the scope of this thesis.

Finally, a 2016 study by Pinet, Ziegler and Alario had aims similar to Berg’s (2002) study and explored the similarities between the writing and typing processes on the psycholinguistic level. However, it did not rely on errors in the two systems to draw the comparison, and both typing and writing errors were excluded from the analysis. The conclusion of the study echoed Berg’s in that writing and typing were seen as sharing the same processes on the psycholinguistic level.

The following chapters introduce the corpus of typing errors created for the needs of this thesis and the accompanying classification, which heavily relies on that by Berg (2002). The main goal of this thesis is to compare this newly collected data with the already available data on typing errors in hope of either confirming the results and observations shown in the cited studies or, more interestingly, offering different ones.

3. METHODOLOGY

The process of collecting data for analysis was certainly a complex one, as it consisted of many steps and relied on two separate tools to collect all the necessary data. The following sections deal with the tools in question and the steps the participants were asked to complete in order to build the required dataset.

3.1. The typing test

The typing test and its execution were arguably the most important part of the data collection. More specifically, the most important part was ensuring that the tool used offered the ability to collect and analyze what Baba and Suzuki (2012) call "corrected errors" – those "errors that are corrected on the spot" (p. 373) by the typist and can only be analyzed either by direct observation of the participant as they type or by using a keylogging tool. One such freely and widely available tool featuring both typing and keylogging functionality was *Key Hero*, a website offering customizable typing tests.

The text that was chosen was a 364-word long excerpt from the English translation of Franz Kafka's novella "Metamorphosis" accessed via the Project Gutenberg online library. The excerpt in question (see Appendix A) was deemed to contain mostly simple vocabulary which would not be an issue for most participants and as such would not interfere with the pace and flow of their typing. The excerpt was then split into two separate typing tests, the first one, titled "MetamorphosisExcerpt [1/2]", containing 165 words and the second one, titled "MetamorphosisExcerpt [2/2]", containing the remaining 199 words[8]. After completing each of the tests, the participants were asked to screen capture the results of their test and upload the image later in the survey. These images contain all the corrected and uncorrected errors the participants made while typing, as well as metrics on their typing speed (expressed in words per minute (WPM)) and accuracy (expressed in percentages). One such image is shown in Figure 1.

---

[8] The reason for the division is purely functional, as *Key Hero* struggles with displaying larger texts, which then results in the participant not being able to see the text as they are typing.

Figure 1

Typing Test Results Containing Corrected and Uncorrected Error Data and Additional Metrics

The images were then transcribed using a word-processing software, *Key Hero*'s built-in classification and color-coding system included. This system is shown in Figure 2.

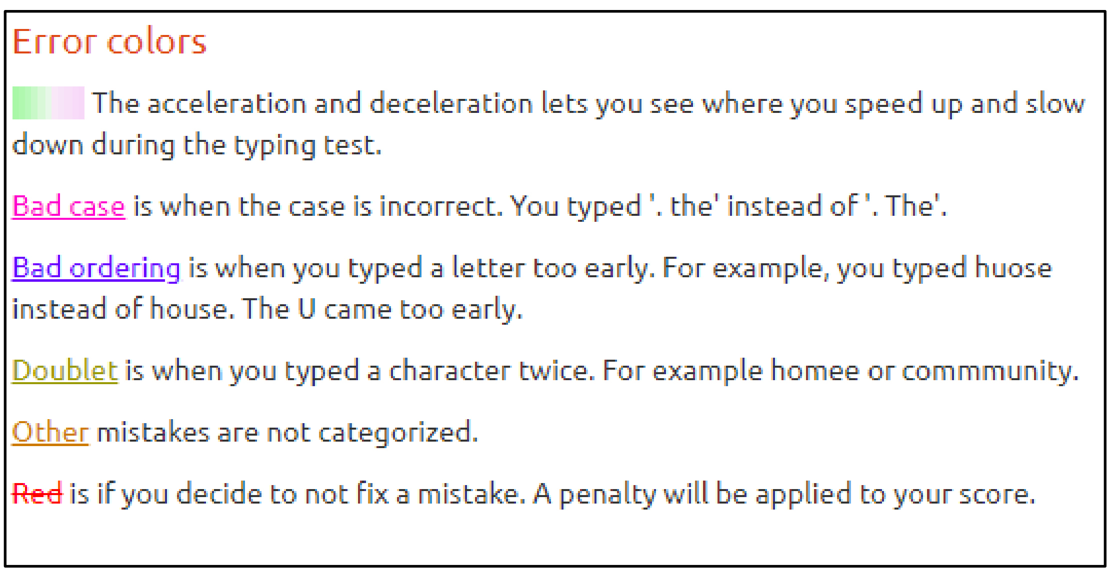

Figure 2

Key Hero*'s Error Classification and Color-Coding*

Key Hero was an invaluable tool in the early stages of the study. It offered almost all of the necessary functions and metrics, while still being simple and accessible enough for all users independent of their computer literacy skills. It did, however, introduce some challenges too. These challenges, as well as those resulting from the very nature of this kind of study, are briefly discussed in the Challenges section of this chapter.

3.2. The survey

Upon finishing the typing tests and screen capturing the results for each half of the text, the participants were asked to fill out a survey titled "Exit survey" (see Appendix B) via Google Forms. The survey contains some basic demographic data about the participants: their gender, age, country or state of residence, level of education and a self-assessment of their typing skill on a scale from 1 to 5 (1 being bad and 5 being excellent). A parameter relating to the keyboard layout used by each participant was added upon receiving feedback from an early participant. Finally, the participants were asked to upload the pictures of their test results for analysis.

3.3. The participants

Twenty-five native speakers of English participated in the study. The participants were aged between 17 and 42 years old (M=27.56) and were predominantly women (68%). Their places of residence varied, but most participants came either from the USA (>40%) or South Africa (>25%). The remaining participants came from other English-speaking countries, with the exception of two participants residing in Denmark and the Philippines. The participants' level of education ranged from high school to master degrees, with the majority of participants having attained a bachelor's degree. Participant data regarding gender, place of residence, and level of education are shown in Figures 3, 4, and 5, respectively. Complete participant data are provided in Appendix C.

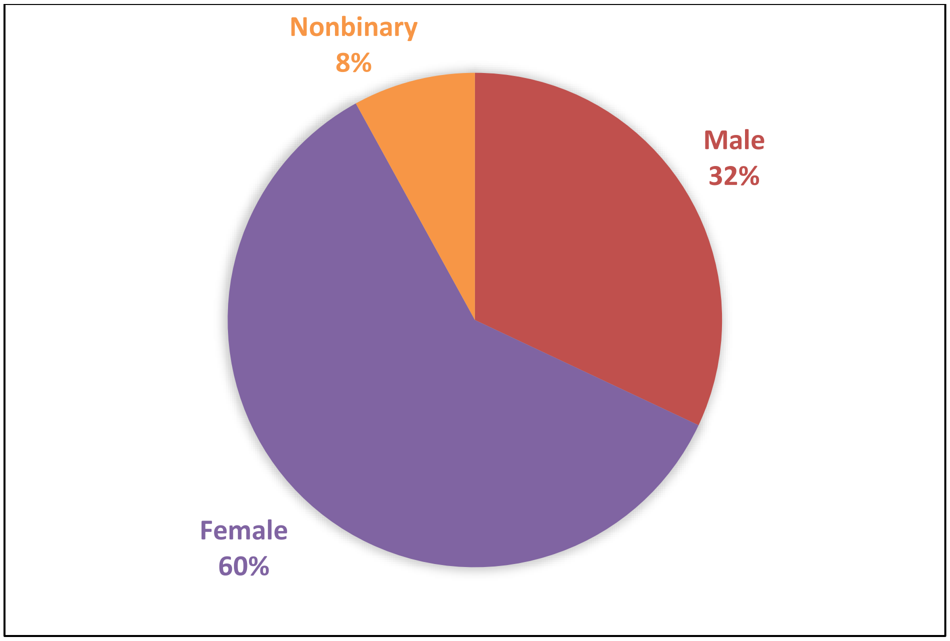

Figure 3

Gender

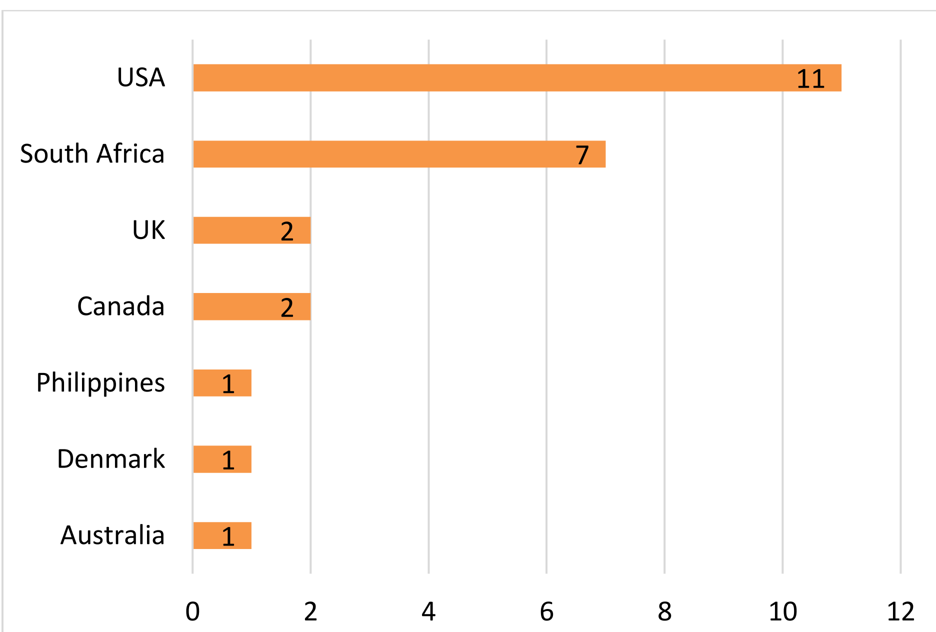

Figure 4

Place of Residence

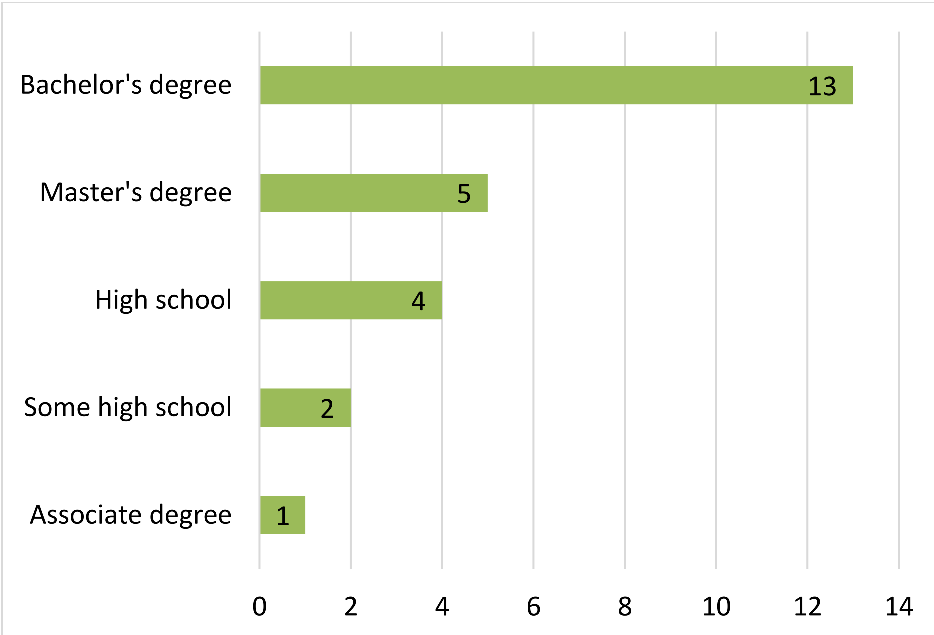

Figure 5

Level of Education

3.4. Challenges

As with any study, there were many challenges and obstacles to overcome before the desired outcome was achieved. Some of them were as simple as struggling to attract enough participants or being inconvenienced by *Key Hero*'s lack of export/share option. However, some obstacles were much more significant. As mentioned before, *Key Hero* has proven to be immensely helpful, but it also introduced some major problems that affected both the quality and the quantity of the collected data.

The lesser of the problems stems from the fact that *Key Hero* makes error detection somewhat too easy for the participant – as soon as they make a wrong keystroke, the entire word gets highlighted in red and stays that way until the participant either corrects the mistake or moves on to the next word. The assumption is that this would severely limit or even completely eliminate the possibility of observing and analyzing the factors that lead to error detection. This is a lesser problem simply because the aim of this thesis is to compare this dataset to Berg's (2002) corpus which lacks any data on error detection rates because it is a corpus of uncorrected typing errors. Therefore, assuming that this functionality doesn't significantly reduce error production, the impact on the dataset is limited. The other problem, however, is potentially more impactful. *Key*

Hero can reliably track all keystrokes and backspaces, which it also uses to determine and categorize all strings of text that do not match the target text. And while it also reliably displays all keystrokes in the error overview, it does not in any way display or mark backspaces, meaning that there is no way to retroactively determine each and every keystroke that was made during typing. This issue was especially noticeable in two different scenarios. The first is in cases of corrected errors where two or more errors occurred consecutively, as in example (1), so that an error was made, then erased, and then immediately after a different error was made in the same location.

1. slightest short**vocoming**tcoming / for: slightest shortcoming

What seemingly happened in example (1) is that the participant typed "short*v*o-" introducing a motor error by substituting the source unit 'c' with the error unit 'v'. At this point, they pressed the backspace key three times instead of two (most likely without noticing) reducing the string to "shor-" and continued typing the string from the point where the original error occurred. This resulted in the second error occurring in the form of an omitted 't', creating the string "shor_coming". The only possible way to arrive at this conclusion is by observing the surrounding letters not marked as errors by *Key Hero*. We see that the string was correct until the 'v' for 'c' motor error, meaning that "short-" was already typed out. But the unmarked string on the other end of the error is "-tcoming", suddenly containing a 't' that was already typed out previously. The conclusion, then, was that in correcting one error another one was made. However, not all such cases could be resolved this way and as a result some were either discarded or were classified based on what could be observed with no guarantee of accuracy. The second scenario which was observed and could potentially have impacted the dataset was that of errors marked as uncorrected. This *Key Hero* feature works in the following way: if the participant makes a mistake and does not correct it either while still typing the word in which the error occurs or while typing the word following it, the error is marked in red and crossed through. The issue, however, is that even if the participant does detect the error and goes back to correct it, the error overview still shows the ultimately corrected segments as uncorrected, having no way to distinguish between eventually-corrected errors and uncorrected ones. This might have resulted in a skewed uncorrected error corpus as it might feature errors which should really be considered corrected. However, as there is no way to discern between the cases of accurately and inaccurately marked uncorrected errors, the decision was made to adhere to *Key Hero*'s classification so as to still try to separate uncorrected errors from corrected ones.

The last notable challenge stems from the fact that the corpus in question is almost entirely made up of corrected errors. Considering that corrected errors often get corrected almost immediately after they occur, it can prove quite difficult to determine what kind of error actually took place. Example (2) illustrates this issue very well:

2. too not**e**iceable / for: too noticeable

The participant substituted an 'i' with an 'e', stopped immediately and erased the erroneous 'e', and finally typed the rest of the word correctly. Seeing this corrected error, at least two possible interpretations can be given: (i) the participant intended to type "not*e*ceable", which would classify the error as a substitution, or (ii) the participant intended to type "not*e*iceable", which would classify the error as an addition. There is no real solution to such cases, but steps were taken to try and narrow down possible interpretations, i.e. all such cases were resolved on individual basis.

3.5. Data and classification

The previously described steps produced a large quantity of raw data that required processing, categorization, and analysis. As mentioned earlier, images submitted by the participants had to be transcribed and tagged according to *Key Hero*'s color-coded system of error tagging. The errors were then isolated and checked to make sure they corresponded to the original images. The final count was a total of 1195 errors according to *Key Hero*'s classification, slightly higher than the number of total errors according to the standard classification described later in this section. The distribution of errors across categories is presented in Table 11.

Table 11

Frequency of Errors Across Categories in the Key Hero *Classification*

Error category	Number of errors
Bad case	100 (8.3%)
Bad ordering	289 (24.2%)
Doublet	45 (3.8%)
Other	686 (57.4%)
Uncorrected	75 (6.3)
Total	1,195

This discrepancy between the two classification systems is mainly explained by their innate differences, but also by the fact that some errors were discarded because they either could not be explained using the selected classification or were a result of external factors (i.e., hardware issues). Example (3) demonstrates the former, and example (4) the latter case:

3. business? **? Was Q was Was it really** / for: business? Was it really
4. he s**sssdwwang** himself / for: he swang himself

Additionally, all errors containing the '=' symbol were discarded, as the symbol itself does not appear in the text and all its occurrences were a result of what could be classified as a spatial error[9] – the '=' key is located to the left of the backspace key and some participants would press it on accident. So, while these errors could easily be classified with the chosen classification, they do not involve any actual letters or even any keys connected to the word or sentence boundary (e.g., commas, full stops, and empty spaces) and were hence discarded from the final dataset. Finally, a small number of competence errors was detected and discarded.

Having collected and processed the dataset, the next step was selecting the most appropriate classification system, if there was one, or otherwise developing a new one. In this case, a classification system was already available – the one presented by Berg (2002). This system was almost entirely[10] adopted for the needs of this thesis, with two small alterations. The first is a simple expansion of one of the proposed classes – contextual within-word addition. During the data processing, it was observed that two different kinds of contextual within-word addition errors could occur: (i) errors where the source unit would not be adjacent to the error unit, e.g., "prop*o*er" for "proper", and (ii) errors where the source unit would be adjacent to the error unit, effectively resulting in doubling of the source unit, e.g., "proper*r*" for "proper". Therefore, a decision was made to subdivide the contextual within-word addition class into the nonadjacent contextual within-word addition and adjacent contextual within-word addition (or alternatively, monopositional doubling error).

---

[9] In MacNeilage's (1964) classification
[10] The positioning of adjacent switches within the word was not examined.

The second alteration relates to the non-contextual monopositional error class – when discussing such errors, Berg (2002, p. 189) briefly refers to MacNeilage (1964) and his classification, stating that some of the errors that he himself classifies as non-contextual (i.e., those whose source unit cannot be found in their vicinity) could be classified as what MacNeilage (1964) calls the spatial error class (i.e., the class of errors resulting from the pressing of a key adjacent to the target one). However, Berg's standard classification does not utilize the spatial error class, nor does it introduce a new class that would play the same role. We deemed that such a class would be beneficial in separating those non-contextual errors that occur at the executive level of the production mechanism for typing (MacNeilage, 1964, p. 158) from those errors occurring at the programming level of the production mechanism. Therefore, a spatial error class was added as a subclass of non-contextual monopositional errors. Furthermore, the original definition of spatial errors was modified and expanded to include not only non-contextual substitution errors (which is how MacNeilage viewed it), but also non-contextual addition errors, which seemingly occur as a result of accidental simultaneous pressing of two (or more) adjacent keys (i.e., 'fat-fingering'). The distinction between spatially-motivated non-contextual substitutions/additions and those that were not spatially-motivated (meaning that the error originated in the programming mechanism) was made primarily based on the proximity of the target unit (i.e., the key that was intended to be pressed) and the substituted/added unit (i.e., the key that ended up being pressed instead or in addition). If they were adjacent to each other, and no other motivations for the error could be detected, the error was classified as spatial. However, the same could not have been done for non-contextual omissions. Such errors certainly occur, most likely in cases where a string contains clusters of units that are all adjacent to each other on the keyboard (e.g., "was"), but such errors could not be discerned from non-contextual omissions with certainty, so we decided to classify all non-contextual omission errors as just that and not include any of them in the spatial error category.

In addition to the spatial error class, one more error category introduced by MacNeilage was considered – multiple classification category, which contains errors that could be placed into multiple error categories (p. 147). However, as all errors have a certain degree of ambiguity to them, there is a significant portion of errors that could theoretically be classified as two or more categories. We therefore decided to place every such case into one definitive category based on supplementary data – patterns observed in other errors made by the same participant, frequencies of the same error between different participants, error replication probability, etc.

With the classification in place and the error dataset ready, the manual review and categorization process could begin. First, each erroneous string of text had to be formatted in a way that would accurately reflect how and where an error occurred, as well as when it was detected and corrected. The formatting was based on the one Berg (2002) used to mark examples, but it too had to be expanded as he used an uncorrected error corpus, whereas 90% of the dataset in our study consisted of corrected errors. The error-marking format used is fully explained in Appendix D. Example (5) illustrates the adaptation of *Key Hero*'s original format for clear error presentation.

5. "visit**por**or's first words" > "visit*p*or['s first words]"

Afterwards, each error was classified on 3 different levels. The first level categorized errors into two categories: lexical and sublexical. An error was classified as lexical when a different word was used in place of the target one (6), when a word was completely omitted (7), or when a word was displaced to a different location within the sentence (8)). Sublexical errors, on the other hand, were all those that occurred within the target word (9).

6. the visitor's *voice* / for: the visitor's first words
7. turned __ and rubbed it / for: turned it and rubbed it
8. clerk *himself* have to come / for: clerk have to come himself
9. of *ht*e / for: of the

While all lexical errors were marked, categorized, and counted, they were excluded from the analysis as Berg's corpus excluded such errors as well. Frequencies of lexical and sublexical errors (both corrected and uncorrected) are shown in Table 12. All subsequent discussions and analyses refer to sublexical errors only.

Table 12

Frequencies of Corrected and Uncorrected Lexical and Sublexical Typing Errors

	Corrected	Uncorrected	Total
Lexical errors	68 (97.1%)	2 (2.9%)	70
Sublexical errors	854 (90.5%)	90 (9.5%)	944

The division of errors into corrected and uncorrected served as the second level of classification. As previously mentioned, the division was largely based on *Key Hero*'s flawed categorization. Unsurprisingly, the main corpus was mostly comprised of corrected sublexical errors – 854 of them to be exact – whereas the remaining 90 errors were classified as uncorrected sublexical errors. Finally, the third level of classification categorized errors based on the number of error units into monopositional and bipositional. Monopositional errors dominated both the corrected error and the uncorrected error subcorpus, comprising ~87% and ~91% of all errors in the subcorpora, respectively. The exact counts of monopositional and bipositional sublexical errors in both subcorpora are shown in Table 13.

Table 13

Frequencies of Monopositional and Bipositional Sublexical Typing Errors

	Monopositional errors	**Bipositional errors**	**Total**
Corrected subcorpus	744 (87.1%)	110 (12.9%)	854
Uncorrected subcorpus	82 (91.1%)	8 (8.9%)	90

As Berg's analysis did not include any errors "involving punctuation marks and other typographical devices (e.g., capitalization)" (p. 187), these errors had to be identified and excluded from the two subcorpora. Initially, there was no defined classification for these types of errors as they were intended to be discarded outright. However, some interesting errors were noted, which motivated the development of a dedicated set of classes to clearly categorize these kinds of errors. Such errors were roughly divided into (i) errors involving empty spaces (i.e., the result of pressing the spacebar on the keyboard), (ii) errors involving punctuation marks, and (iii) errors in word capitalization. A further subdivision of each main class and the frequencies of each error type are shown in Table 14 for the corrected error subcorpus and in Table 15 for the uncorrected error subcorpus.

Table 14

Empty Space, Punctuation Mark, and Capitalization Errors in the Corrected Error subcorpus

		Count	**Total**
Empty space errors	Empty space adjacent switch	20	98
	Empty space omission	18	
	Empty space addition	60	
Punctuation mark errors	Punctuation mark adjacent switch	5	72
	Punctuation mark substitution	39	
	Punctuation mark omission	18	

	Punctuation mark addition	10	
Capitalization errors			19

Table 15

Empty Space, Punctuation Mark, and Capitalization Errors in the Uncorrected Error Subcorpus

		Count	**Total**
Empty space errors	Empty space adjacent switch	3	5
	Empty space omission	1	
	Empty space addition	1	
Punctuation mark errors	Punctuation mark adjacent switch	0	11
	Punctuation mark substitution	6	
	Punctuation mark omission	4	
	Punctuation mark addition	1	
Capitalization errors			9

From this data, some interesting observations can be drawn. Firstly, there seems to be a large number of empty space additions in the corrected error subcorpus. This appears to be the result of some participants' habit or preference for text formatting – 44 of the 60 detected empty space additions were a result of the participant adding an empty space preceding a punctuation mark, as in (10).

10. they danced around•[.] / for: they danced around.

This error being a result of habit seems to be supported by the fact that all such errors almost exclusively occurred multiple times per participant (but not all participants made such errors) and were rarely a one-time occurrence. The other interesting observation was that bipositional errors involving empty spaces and punctuation marks could and would occur. Such errors were classified as empty space adjacent switches and punctuation mark adjacent switches, respectively. Empty space adjacent switches could either involve a word-final unit (11) or a word-initial one (12). Punctuation mark switches tended to involve a letter and a punctuation mark (13), most often an apostrophe, but there was also an instance of different adjacent punctuation marks switching places.

11. became al·*l*the [more lively] / for: became all the more lively

12. a little by*t*·he carpet / for: a little by the carpet

13. mai*’d*s[firm steps] / for: maid’s firm steps

By removing all such errors from the corrected and the uncorrected sublexical error subcorpora, the error count was reduced to 665 and 65 errors, respectively – we thus obtained the final error count that was classified using the extended standard classification and then analyzed and compared. Monopositional error classes and their frequencies in the corrected error subcorpus are presented in Table 16, followed by the bipositional error classes and their frequencies in the same subcorpus in Table 17. Their uncorrected counterparts are presented in Tables 18 and 19 respectively. The analysis, interpretation, and comparison of the data is then presented in Chapter 4.

Table 16

Monopositional Error Classification and Error Frequency in the Corrected Error Subcorpus

			Count
Contextual within-word errors	Substitution	Anticipatory substitution	15
		Perseveratory substitution	7
	Omission (deletion)	Adjacent masking	10
		Nonadjacent anticipatory masking	4
		Nonadjacent perseveratory masking	31
		Plain contextual omission	0
	Addition	Adjacent addition	54
		Nonadjacent addition	25
Contextual between-word errors	Substitution	Anticipatory substitution	15
		Perseveratory substitution	18
	Omission (deletion)	Anticipatory masking	2
		Perseveratory masking	20
		Plain contextual omission	4
	Addition		28
Non-contextual errors	Substitution		7
	Omission (deletion)		111
	Addition		17
	Spatial (motor) errors		211

Table 17

Bipositional Error Classification and Error Count in the Corrected Error Subcorpus

			Count
Adjacent switch	$C_1C_2 \rightarrow C_2C_1$		19
	$V_1V_2 \rightarrow V_2V_1$		2
	$CV \rightarrow VC$		23
	$VC \rightarrow CV$		24
	Grapheme doubling	Doubling after adjacent switch	0
		Misordered adjacent grapheme doubling	11
Reversal (nonadjacent switch)	Single grapheme		2
	Grapheme doubling	Doubling after reversal	0
		Misordered nonadjacent grapheme doubling	0
Anticipatory shift			4
Perseveratory shift			2

Table 18

Monopositional Error Classification and Error Frequency in the Uncorrected Error Subcorpus

			Count
Contextual within-word errors	Substitution	Anticipatory substitution	2
		Perseveratory substitution	0
	Omission (deletion)	Adjacent masking	0
		Nonadjacent anticipatory masking	1
		Nonadjacent perseveratory masking	2
		Plain contextual omission	0
	Addition	Adjacent addition	7
		Nonadjacent addition	2
Contextual between-word errors	Substitution	Anticipatory substitution	0
		Perseveratory substitution	0
	Omission (deletion)	Anticipatory masking	4
		Perseveratory masking	1
		Plain contextual omission	0
	Addition		2
Non-contextual errors	Substitution		0
	Omission (deletion)		8

	Addition	1
	Spatial (motor) errors	30

Table 19

Bipositional Error Classification and Error Count in the Uncorrected Error Subcorpus

			Count
Adjacent switch	$C_1C_2 \rightarrow C_2C_1$		1
	$V_1V_2 \rightarrow V_2V_1$		1
	CV → VC		3
	VC → CV		0
	Grapheme doubling	Doubling after adjacent switch	0
		Misordered adjacent grapheme doubling	0
Reversal (nonadjacent switch)	Single grapheme		0
	Grapheme doubling	Doubling after reversal	0
		Misordered nonadjacent grapheme doubling	0
Anticipatory shift			0
Perseveratory shift			0

4. RESULTS AND ANALYSIS

The following sections offer a deeper look into the presented data. Error frequencies are discussed and compared amongst themselves and with data from Berg's (2002) study.

4.1. The uncorrected error subcorpus

Despite the clear and total dominance of corrected errors in the main corpus, there is reason to consider the uncorrected error subcorpus in isolation, despite its small size. Firstly, major differences are bound to emerge between the two subcorpora. Secondly, an uncorrected error subcorpus is especially valuable for comparison with Berg's data as his findings are based exclusively on an uncorrected typing error corpus. This comparison does come with some caveats, however. For one, Berg's corpus is significantly larger – 14 times so – than the uncorrected error subcorpus. Berg's corpus also spans 433 sources, compared to the 11 that produced the uncorrected error subcorpus.[11] Furthermore, the 433 sources were articles and individual chapters, all unique, and it may be assumed that such variety in target texts would also produce a more varied error dataset. On the other hand, a lack of variety in target texts might be expected to produce a less varied error dataset, albeit one that features recurring errors across multiple sources. It should also be noted that the errors found in Berg's corpus are those that 'slipped through the cracks' in the process of repeated proofreading and error checking. This is not the case with our corpus of uncorrected errors as these errors would likely have been detected and corrected upon the first re-reading of the text, if not sooner. For this exact reason, all uncorrected spatial errors were excluded from the analysis, as it is likely that Berg's corpus contains only a small number of such errors considering how easy they would have been to detect by their authors during production.

The (sub)corpora comparison was done using the (somewhat limited) data provided by Berg (2002, p. 189–194). All the frequency tables[12] discussed in his study were replicated here, including their original values. A distinction was made between Berg's data, titled in the following tables as *Typing 1 (Berg)*, whereas data pertaining to our uncorrected error subcorpus were titled

---

[11] The other 14 participants contributed no uncorrected mistakes. Of the 11 participants that did produce uncorrected mistakes, one participant produced 60% of all uncorrected errors, another participant produced 21%, and the remaining 19% of the errors was produced by the other 9 participants.

[12] Bar Table 7 (p. 196).

Typing 2 (UC). Table 20 presents the comparison of contextual and non-contextual error frequencies across all modalities. The writing and speaking data used in the comparison are the ones provided by Berg.

Table 20

Frequency of Contextual and Non-contextual Errors in Typing, Writing, and Speaking

	Contextual	Non-contextual
Typing 2 (UC)	21 (70.0%)	9 (30.0%)
Typing 1 (Berg)	359 (71.8%)	141 (28.2%)
Writing	514 (74.1%)	180 (25.9%)
Speaking	1,845 (85.1%)	324 (14.9%)

When comparing the overall frequencies of contextual and non-contextual typing errors, the uncorrected error subcorpus does not show any deviation from the data provided by Berg. If anything, it strengthens his assessment that "the factors that bring about contextual errors have a more important role to play in speaking than in writing and typing" (p. 189). Table 21 contains a comparison of contextual between-word and contextual within-word error frequencies across all modalities.

Table 21

Frequency of Contextual Between- and Within-Word Errors in Typing, Writing, and Speaking

	Between word	Within-word
Typing 2 (UC)	7 (33.3%)	14 (66.7%)
Typing 1 (Berg)	39 (10.9%)	320 (89.1%)
Writing	106 (20.6%)	408 (79.4%)
Speaking	1,605 (87.0%)	240 (13.0%)

At this point, some differences between Berg's uncorrected error corpus and ours can be observed. While contextual within-word errors remain dominant, the gap between the two error types seems to be smaller than is the case for Berg's corpus and the corpus of slips of the pen. Table 22 presents a comparison of substitution, addition, and omission error frequencies across all modalities.

Table 22

Frequency of Substitution, Addition, and Omission Errors in Typing, Writing, and Speaking

	Substitutions	Additions	Omissions
Typing 2 (UC)	2 (6.7%)	12 (40%)	16 (53.3%)
Typing 1 (Berg)	217 (43.4%)	83 (16.6%)	200 (40.0%)
Writing	245 (45.5%)	95 (17.6%)	199 (36.9%)
Speaking	2,249 (88.0%)	200 (7.8%)	107 (4.2%)

The proportions of substitutions, additions, and omissions in our uncorrected error subcorpus are distinctive compared to the other two modalities and to Berg's corpus. Omission frequencies are mostly in line with other modalities, but appear to be dominant only in our subcorpus. Additions and substitutions, however, exhibit inverse proportions compared to Berg's corpus and the corpus of slips of the pen. This discrepancy might be due to the previously described fundamental differences between our subcorpus and Berg's, or even the impact of human judgement in error categorization. Table 23 contains a comparison of non-contextual omissions, masking errors[13], and contextual omissions across all modalities.

Table 23

Frequency of Non-contextual Omissions, Masking Errors, and Contextual Omissions in Typing, Writing, and Speaking

	Non-contextual omissions	Masking errors	Contextual omissions
Typing 2 (UC)	8 (50. 0%)	8 (50. 0%)	0 (0.0%)
Typing 1 (Berg)	76 (38.2%)	123 (61.8%)	0 (0.0%)
Writing	108 (54.3%)	91 (45.7%)	0 (0.0%)
Speaking	26 (23.0%)	38 (33.6%)	49 (43.4%)

Omission errors seem to mostly align with Berg's findings. Contextual omissions are completely absent, which is also in line with Berg's corpus and the corpus of slips of the pen. The frequencies of non-contextual omissions and masking errors in the uncorrected error subcorpus

[13] Masking errors were introduced in Berg (2002) and in Table 6 they are defined as a subtype of contextual omission errors, which refer to an error resulting from one element blotting out an identical unit in the vicinity.

are more similar to those of the writing corpus than that of Berg's, but the difference is slight. Table 24 shows a comparison of directionality of contextual between-word errors across all modalities.

Table 24

Directionality of Contextual Between-Word Errors in Typing, Writing, and Speaking

	Masking		**Substitution**	
	Anticipation	**Perseveration**	**Anticipation**	**Perseveration**
Typing 2 (UC)	4 (80.0%)	1 (20.0%)	0 (0.0%)	0 (0.0%)
Typing 1 (Berg)	42 (43.3%)	55 (56.7%)	37 (54.4%)	31 (45.6%)
Writing	28 (41.8%)	39 (58.2%)	111 (65.3%)	59 (34.7%)
Speaking	38 (45.8%)	45 (54.2%)	839 (62.7%)	500 (37.3%)

The data on masking and substitution directionality in between-word errors shows the greatest deviations of all comparisons so far. Between-word substitution errors are completely absent from the uncorrected error corpus, whereas all other modalities exhibit fairly balanced proportions of both anticipatory substitutions and perseveratory ones. While masking errors do occur in the subcorpus, their proportions are drastically different from those in other modalities, with anticipatory masking errors being four times more common than perseveratory ones. Table 25 presents a comparison of adjacent switch error frequencies across all modalities.

Table 25

Frequency of Adjacent Switches in Typing, Writing, and Speaking

	C_1C_2 → C_2C_1	**V_1V_2 → V_2V_1**	**CV → VC**	**VC → CV**	**Totals**
Typing 2 (UC)	1 (20.0%)	1 (20.0%)	3 (60.0%)	0 (0.0%)	5
Typing 1 (Berg)	14 (14.1%)	17 (17.1%)	40 (40.4%)	28 (28.3%)	99
Writing					
Complete	5 (20.0%)	3 (12.0%)	7 (28.0%)	10 (40.0%)	25
Incomplete	29 (24.2%)	17 (14.2%)	43 (35.8%)	31 (25.8%)	120
Speaking	6 (60.0%)	0 (0.0%)	1 (10.0%)	3 (30.0%)	10

Adjacent switches again align with the comparison corpora, for the most part. One obvious exception is the absence of VC → CV adjacent switches in our uncorrected error subcorpus. This category, while somewhat less frequent in Berg's corpus, is still present in all other corpora. And just like in Berg's corpus, CV → VC adjacent switches occur most frequently. The comparison between our uncorrected error subcorpus and Berg's corpus reveals that the proportions are largely consistent with both that corpus and the written corpus. The similarities are the strongest when comparing major error categories and they become more distinct as the categories become more specific.

Due to its limited size, the uncorrected error subcorpus has limited potential for analysis beyond its comparison with other corpora. The most commonly observed error type was non-contextual omission (14), accounting for 23% of all errors with a total of 8 instances, followed by adjacent contextual within-word addition (13) with 7 instances (20%).

14. all h_s force / for: all his force
15. investigate*e* it / for: investigate it

The complete uncorrected error subcorpus can be found in Appendix E.

4.2. The corrected error subcorpus

Unlike the uncorrected error subcorpus, the corrected error subcorpus is better suited for comparison with Berg's corpus and other modalities. This is mainly due to its more substantial size – even with the surprisingly numerous spatial errors (211) excluded from the comparison, 454 errors remain. The comparison from the previous section was replicated in this one as well, with the addition of the corrected error subcorpus data obtained in this research, which was titled *Typing 3 (C)* in the tables. Table 26 shows a comparison of contextual and non-contextual error frequencies across all modalities.

Table 26

Frequency of Contextual and Non-contextual Errors in Typing, Writing, and Speaking

	Contextual	**Non-contextual**
Typing 3 (C)	233 (63.3%)	135 (36.7%)
Typing 2 (UC)	21 (70.0%)	9 (30.0%)
Typing 1 (Berg)	359 (71.8%)	141 (28.2%)
Writing	514 (74.1%)	180 (25.9%)
Speaking	1,845 (85.1%)	324 (14.9%)

Comparing the frequencies of contextual and non-contextual typing errors reveals that the corrected error subcorpus shows only a slight deviation from the data provided by Berg. Just like the uncorrected error subcorpus, it too strengthens Berg's assessment that "the factors that bring about contextual errors have a more important role to play in speaking than in writing and typing" (p. 189). Table 27 presents a comparison of between-word and within-word error frequencies across all modalities.

Table 27

Frequency of Between- and Within-Word Errors in Typing, Writing, and Speaking

	Between word	Within-word
Typing 3 (C)	87 (37.3%)	146 (62.7%)
Typing 2 (UC)	7 (33.3%)	14 (66.7%)
Typing 1 (Berg)	39 (10.9%)	320 (89.1%)
Writing	106 (20.6%)	408 (79.4%)
Speaking	1,605 (87.0%)	240 (13.0%)

The corrected error subcorpus follows the trend established by the uncorrected error subcorpus. Within-word errors are still clearly more common, but to a lesser degree when compared to Berg's corpus and the corpus of slips of the pen. Table 28 shows a comparison of substitution, addition, and omission frequencies across all modalities.

Table 28

Frequency of Substitutions, Additions, and Omissions in Typing, Writing, and Speaking

	Substitutions	Additions	Omissions
Typing 3 (C)	62 (16.8%)	124 (33.7%)	182 (49.5%)
Typing 2 (UC)	2 (6.7%)	12 (40%)	16 (53.3%)
Typing 1 (Berg)	217 (43.4%)	83 (16.6%)	200 (40.0%)
Writing	245 (45.5%)	95 (17.6%)	199 (36.9%)
Speaking	2,249 (88.0%)	200 (7.8%)	107 (4.2%)

Once again, the corrected error subcorpus mirrors the patterns seen in the uncorrected error subcorpus, with slightly diluted proportions. Omissions remain the most prevalent error class, followed by additions and substitutions. However, the difference between each category is less pronounced than in the uncorrected error subcorpus. Still, the inverse relationship between substitutions and additions compared to Berg's corpus and the corpus of slips of the pen remains unchanged. Considering the corrected nature of the subcorpus, this peculiar ratio of substitutions and additions could be due to the fact that many of these errors had to be classified without knowing the full context – most errors were detected as soon as the erroneous unit was produced, which could have skewed the results towards additions or omissions. Table 29 presents a comparison of non-contextual omission, masking, and contextual omission error frequencies across all modalities.

Table 29

Frequency of Non-contextual Omissions, Masking Errors, and Contextual Omissions in Typing, Writing, and Speaking

	Non-contextual omissions	Masking errors	Contextual omissions
Typing 3 (C)	111 (61.0%)	67 (36.8%)	4 (2.2%)
Typing 2 (UC)	8 (50. 0%)	8 (50. 0%)	0 (0.0%)
Typing 1 (Berg)	76 (38.2%)	123 (61.8%)	0 (0.0%)
Writing	108 (54.3%)	91 (45.7%)	0 (0.0%)
Speaking	26 (23.0%)	38 (33.6%)	49 (43.4%)

The corrected error subcorpus has a small number of contextual omissions, thus slightly deviating from the complete absence of contextual omissions observed in the uncorrected subcorpus, Berg's corpus and the written corpus. The 50:50 ratio between masking errors and non-contextual omissions in the uncorrected error corpus is not reflected in the corrected error subcorpus, with non-contextual omissions accounting for just over 60% of all omission errors, and masking errors close to 37%. In Berg's corpus this relationship is once again inverse, but interestingly, the same cannot be said for the corpus of slips of the pen, where we observe almost an even split, with non-contextual omissions being somewhat more common of the two. Table 30 shows a comparison of directionality of contextual between-word errors across all modalities.

Table 30

Directionality of Contextual Between-Word Errors in Typing, Writing, and Speaking

	Masking		Substitution	
	Anticipation	**Perseveration**	**Anticipation**	**Perseveration**
Typing 3 (C)	2 (9.0%)	20 (91.0%)	15 (45.5%)	18 (54.5%)
Typing 2 (UC)	4 (80.0%)	1 (20.0%)	0 (0.0%)	0 (0.0%)
Typing 1 (Berg)	42 (43.3%)	55 (56.7%)	37 (54.4%)	31 (45.6%)
Writing	28 (41.8%)	39 (58.2%)	111 (65.3%)	59 (34.7%)
Speaking	38 (45.8%)	45 (54.2%)	839 (62.7%)	500 (37.3%)

Directionality of between-word errors once again reveals a compelling deviation of both subcorpora from the comparison data. The uncorrected error subcorpus displayed two oddities. First, a complete absence of any between-word substitution errors. This is not replicated in the corrected error corpus – on the contrary, anticipatory and perseveratory between-word substitution error frequencies mostly match those in Berg's corpus. The second oddity of the uncorrected error subcorpus was the high frequency of anticipatory between-word masking errors. Not only is this not the case in the corrected error subcorpus, but the opposite is actually true – anticipatory between-word masking errors now account for less than 10% of all between-word masking errors. Once again, the most likely explanation for this discrepancy lies in the categorization process and the challenges of categorizing errors without sufficient context. Table 31 presents a comparison of adjacent switches frequencies across all modalities.

Table 31

Frequency of Adjacent Switches in Typing, Writing, and Speaking

	$C_1C_2 \rightarrow C_2C_1$	$V_1V_2 \rightarrow V_2V_1$	CV → VC	VC → CV	Totals
Typing 3 (C)	19 (28.0%)	2 (2.9%)	23 (33.8%)	24 (35.3%)	68
Typing 2 (UC)	1 (20.0%)	1 (20.0%)	3 (60.0%)	0 (0.0%)	5
Typing 1 (Berg)	14 (14.1%)	17 (17.1%)	40 (40.4%)	28 (28.3%)	99
Writing					
Complete	5 (20.0%)	3 (12.0%)	7 (28.0%)	10 (40.0%)	25
Incomplete	29 (24.2%)	17 (14.2%)	43 (35.8%)	31 (25.8%)	120
Speaking	6 (60.0%)	0 (0.0%)	1 (10.0%)	3 (30.0%)	10

Adjacent switch frequencies in the corrected error subcorpus match the comparison data better than the uncorrected error subcorpus did. Interestingly, while all frequencies between the corrected error subcorpus and Berg's corpus have similar values, they are 'misordered' – in Berg's corpus, CV → VC switch is the most frequent, followed by VC → CV switch, $V_1V_2 \rightarrow V_2V_1$ switch, and finally $C_1C_2 \rightarrow C_2C_1$ switch. In the corrected error subcorpus, the most frequent switch is the VC → CV, followed by the CV → VC switch, then $C_1C_2 \rightarrow C_2C_1$, and finally the $V_1V_2 \rightarrow V_2V_1$ switch. Interestingly, completed adjacent switches in writing follow the same pattern. The corrected error subcorpus mostly aligns with data and proportions found in Berg's corpus, and even more so than the uncorrected error subcorpus, an observation that is especially interesting considering that Berg's corpus was comprised of uncorrected errors.

Compared to the uncorrected error subcorpus, the corrected error subcorpus presents more compelling data and provides more informative examples. By revisiting the tables 16 and 17 from previous sections, percentage-based frequencies for all major error categories can be analyzed and their proportions discussed. Table 32 contains data on frequencies of contextual within-word errors in the corrected error subcorpus.

Table 32

Frequencies of Contextual Within-Word Errors

			Count	Totals
Contextual within-word errors	Substitution	Anticipatory substitution	15 (10.3%)	22 (15.1%)
		Perseveratory substitution	7 (4.8%)	
	Omission (deletion)	Adjacent masking	10 (6.9%)	45 (30.8%)
		Nonadjacent anticipatory masking	4 (2.7%)	
		Nonadjacent perseveratory masking	31 (21.2%)	
		Plain contextual omission	0 (0.0%)	
	Addition	Adjacent addition	54 (37.0%)	79 (54.1%)
		Nonadjacent addition	25 (17.1%)	

By far, the most dominant within-word error category was adjacent addition or doubling. All within-word additions accounted for more than 54% of all contextual within-word errors, an unusual occurrence that was already observed in Table 28 (albeit for all additions across both contextual categories). Adjacent additions in particular are quite interesting as it appears that some of them could be classified as spatial errors, maybe occurring as a result of a slight muscle spasm or accidental increase in applied pressure, and not as a result of erroneous programming. Examples (16) and (17) show two cases of adjacent contextual within-word additions.

16. not enoug*g*h / for: not enough
17. swang*g* himself / for: swang himself

In both examples, the words containing the addition were fully typed, ensuring that it truly was a case of addition, and not a substitution, omission, or bipositional error. Example (16) could be theorized to really be a spatial (addition) error due to the ‘g’ and ‘h’ keys being adjacent to each other and their corresponding letters occurring in a sequence in the target word, but contextually motivated addition cannot be excluded. Example (17) is particularly interesting as it was one of the more frequent errors across participants – 4 out of 25 participants produced this identical addition error. The cause of the error is unclear, but it's relatively high occurrence rate across participants might point to it being contextually motivated rather than not.

If we assumed that all or most adjacent contextual within-word additions are really spatial errors and excluded them from the analysis, the proportions of within-word categories would shift. For the adjusted frequency data, see table 33.

Table 33

Frequencies of Contextual Within-Word Errors Excluding Adjacent Addition Errors

			Count	Totals
Contextual within-word errors	Substitution	Anticipatory substitution	15 (16.3%)	22 (23.9%)
		Perseveratory substitution	7 (7.6%)	
	Omission (deletion)	Adjacent masking	10 (10.9%)	45 (48.9%)
		Nonadjacent anticipatory masking	4 (4.3%)	
		Nonadjacent perseveratory masking	31 (33.7%)	
		Plain contextual omission	0 (0.0%)	
	Addition			25 (27.2%)

Now the frequencies of within-word additions (and additions in general) come closer to those observed by Berg and those that appear in writing. Substitutions, however, still seem to be underrepresented compared to other modalities. One possible conclusion to be drawn is that substitution errors generally occur less frequently but are more difficult to detect. This could possibly explain why they would occur at a higher rate in Berg's corpus but are much less common in the corrected error subcorpus. Example (18) is an anticipatory within-word substitution taken from the subcorpus, whereas (19) exemplifies a perseveratory within-word substitution.

18. *q*nquiries / for: enquiries
19. open*en*[g] / for: opening

Expanded frequency data for between-word errors is presented in Table 34.

Table 34

Frequencies of Contextual Between-Word Errors

			Count	Totals
Contextual between-word errors	Substitution	Anticipatory substitution	15 (17.2%)	33 (38.0%)
		Perseveratory substitution	18 (20.7%)	
	Omission (deletion)	Anticipatory masking	2 (2.3%)	26 (29.9%)
		Perseveratory masking	20 (23.0%)	
		Plain contextual omission	4 (4.6%)	
	Addition			28 (32.2%)

Between-word error frequencies are notably different from within-word ones. While substitutions are technically most common, all three categories occur at similar frequencies. Both in isolation and in comparison with other modalities, omissions seem to be particularly interesting when discussing between-word errors. Firstly, all plain contextual omission errors that can be found in the subcorpus belong to the between-word category of errors. The four plain contextual omission are considered uncommon compared to some other error categories, but they are certainly more common compared to the complete absence of such errors in Berg's corpus and in writing. It could be argued that some of the four cases could be categorized in other ways as well, but some cases seem to fall squarely within the plain contextual omission category, as illustrated by (20) and (21).

20. then ther_ / for: then there
21. at lea_t / for: at least

Example (20) is interesting as it could also be categorized as a nonadjacent perseveratory contextual within-word omission (or even between-word, but word boundaries matter). However, it was categorized as a plain contextual omission because of the strong vowel-consonant structure similarity between the preceding word and the target word where the error occurred. This is further illustrated in example (22).

22. 'then' – 'there' / 'CCVC' – 'CCVCV' → 'then' – 'ther' / 'CCVC' – 'CCVC'

The second noteworthy aspect of between-word omissions is the previously established 1:10 ratio of anticipatory to perseveratory masking errors. A further inspection of all perseveratory masking errors does not reveal the possibility of uncertain categorization or even directionality overlap. It is, however, possible that some errors categorized as between-word masking might be interpreted as non-contextual omissions, but such cases are rare and unlikely to significantly affect the directionality proportions. Examples (23) and (24) are two cases of anticipatory masking, whereas examples (25) and (26) show cases of perseveratory masking.

23. wa_[softened] / for: was softened
24. the _hief clerk / for: the chief clerk
25. that _his was / for: that this was
26. the c_ief / for: the chief

Moving on to addition errors, they once again provide some compelling examples. Like the previously discussed cases of the 'swang' error, example (27) is a between-word addition error, which occurred multiple times across different participants.

27. have the wisdom*e* / for: have the wisdom

Initially this error was classified as a non-contextual addition as it was estimated that the pressure from the 'e' unit in 'the' was not likely to motivate the addition. However, the same error was produced four times by four different participants, and so it was reclassified as a contextual between-word addition error seemingly motivated by the word-final 'e' unit in both 'have' and 'the' preceding the target word. Alternatively, it could be a competence error, but the fact that all of the errors were corrected immediately and were not repeated (neither of which was the case with other competence errors found in the subcorpus) doesn't seem to support that explanation. Another two similar cases of between-word addition occurred in the subcorpus, with two incidences each:

28. couple of hours*e* / for: couple of hours
29. trainees make*s* enquiries / for: trainees make enquiries

Example (28) seems to mimic the process that took place in example (27) – word-final 'e' in 'couple' motivates the addition of a word-final 'e' in 'hours'. It would also seem that the preposition 'of' positioned between 'couple' and 'hours' does not affect this process, possibly due to its length and function. Example (29) again features word-final source units ('s') and a word-final error unit. In this case, the source units both precede and follow the error unit, making it unclear which of the source units is the “real” one.

Finally, Table 35 presents expanded bipositional error data.

Table 35

Frequencies of Bipositional Errors

<table>
<tr><th></th><th></th><th></th><th>Count</th><th>Totals</th></tr>
<tr><td rowspan="6">Adjacent switch</td><td colspan="2">$C_1C_2 \rightarrow C_2C_1$</td><td>19 (21.8%)</td><td rowspan="6">79 (90.8%)</td></tr>
<tr><td colspan="2">$V_1V_2 \rightarrow V_2V_1$</td><td>2 (2.3%)</td></tr>
<tr><td colspan="2">$CV \rightarrow VC$</td><td>23 (26.4%)</td></tr>
<tr><td colspan="2">$VC \rightarrow CV$</td><td>24 (27.6%)</td></tr>
<tr><td rowspan="2">Grapheme doubling</td><td>Doubling after adjacent switch</td><td>0 (0.0%)</td></tr>
<tr><td>Misordered adjacent grapheme doubling</td><td>11 (12.6%)</td></tr>
<tr><td rowspan="3">Reversal (nonadjacent switch)</td><td colspan="2">Single grapheme</td><td>2 (2.3%)</td><td rowspan="3">2 (2.3%)</td></tr>
<tr><td rowspan="2">Grapheme doubling</td><td>Doubling after reversal</td><td>0 (0.0%)</td></tr>
<tr><td>Misordered nonadjacent grapheme doubling</td><td>0 (0.0%)</td></tr>
<tr><td colspan="4">Anticipatory shift</td><td>4 (4.6%)</td></tr>
<tr><td colspan="4">Perseveratory shift</td><td>2 (2.3%)</td></tr>
</table>

What is immediately obvious is the complete absence of all doubling errors outside of the misordered adjacent grapheme doubling which occurred relatively frequently. Some of these errors are shown in (30) and (31).

30. the d*orr* / for: the door

31. His f*aal* / for: His fall

The most frequent bipositional errors by far were adjacent switches in general. Proportions of different adjacent switch combinations are presented in Table 36.

Table 36

Frequencies of Adjacent Switches

<table>
<tr><th></th><th></th><th>Count</th></tr>
<tr><td rowspan="2">Adjacent switch</td><td>$C_1C_2 \rightarrow C_2C_1$</td><td>19 (24.1%)</td></tr>
<tr><td>$V_1V_2 \rightarrow V_2V_1$</td><td>2 (2.5%)</td></tr>
</table>

CV → VC		23 (29.1%)
VC → CV		24 (30.4%)
Grapheme doubling	Doubling after adjacent switch	0 (0.0%)
	Misordered adjacent grapheme doubling	11 (13.9%)

Among the adjacent switches, the VC → CV switches were most common, as in (32), but only slightly more so than the CV → VC switches (33). The third most common switches were $C_1C_2 \rightarrow C_2C_1$ switches (34), and the least common by far were $V_1V_2 \rightarrow V_2V_1$ switches (35). Nonadjacent switches and shifts were very uncommon.

32. a comp*na*y / for: a company
33. investi*ag*te / for: investigate
34. a ri*gn* / for: a ring
35. bec*ua*se / for: because

The complete corrected error subcorpus can be found in Appendix E.

5. CONCLUSION

This study aimed to compare typing and writing in terms of error production. The comparison went beyond just comparing the writing error data with the newly collected typing error subcorpora, as it also incorporated typing error data from Berg (2002). The results suggest that when disregarding errors unique to the keyboard output modality, typing mostly mirrors writing in regards to error production, confirming previous findings by Berg. Some more significant discrepancies between the obtained subcorpora and the available data were observed. The causes of the discrepancies are unclear and could be attributed to various factors such as corpus bias, human judgement, and source material restrictions. Our assessment suggests that differences in corpus type and scope are likely to be the cause of most large discrepancies, while occasional overlap and ambiguity of error categories and produced errors could lead to slight differences in data between different researchers. This study provides valuable data with a lot of potential for future research, including exploring position frequency trends in adjacent switches and relationships between participant demographics and their error production.

6. ABSTRACT

This thesis presents an analysis of typing errors provided by 25 native English-speaking participants. The participants were asked to complete a typing test using the online tool *Key Hero*, which resulted in 25 text units, each containing 364 words. The texts contained data on errors made by the participants, both corrected and uncorrected, which were then isolated and compiled into a typing error corpus. The corpus was further subdivided into the corrected and the uncorrected typing error subcorpora. Errors in each subcorpora were classified using the classification presented in Thomas Berg's 2002 article "Slips of the typewriter key", and the classified error data compared to the data presented by Berg in hopes of replicating the results of his study. The results of the thesis partially matched those of Berg's, with most discrepancies being attributed to corpus bias, human judgement, and source material restrictions. The conclusion reiterated Berg's, indicating that typing most closely resembles writing when it comes to error production.

Keywords: typing errors, slips of the keyboard, uncorrected typing errors, corrected typing errors, typing error corpus, typing error dataset

7. REFERENCES

Baba, Y. and Suzuki, H. (2012). How Are Spelling Errors Generated and Corrected? A Study of Corrected and Uncorrected Spelling Errors Using Keystroke Logs. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 373–377). Jeju Island, Korea: Association for Computational Linguistics. https://aclanthology.org/P12-2073

Berg, T. (2002). Slips of the typewriter key. *Applied Psycholinguistics*, 23(2), 185-207. https://doi.org/10.1017/S0142716402002023

Fromkin, V. A. (1973, December). Slips of the Tongue. *Scientific American*, *229*(6), 110–117. https://doi.org/10.1038/scientificamerican1273-110

Grudin, J. T. (1983). Error patterns in novice and skilled transcription typing. In W. E. Cooper (Ed.), *Cognitive aspects of skilled typewriting* (pp. 121–143). New York: Springer.

Harley, T. A. (2001). *The Psychology of Language From Data to Theory* (2nd ed.). Psychology Press Ltd.

MacKay, D. G. (1993). Slips of the pen, tongue, and typewriter: A contrastive analysis. In G. Blanken, J. Dittmann, H. Grimm, J. Marshall, and C.-W. Wallesch (Eds.), *Linguistic disorders and pathologies* (pp. 66–72). Berlin: de Gruyter.

MacNeilage, P. F. (1964). Typing errors as clues to serial ordering mechanisms in language behaviour. *Language and Speech, 7*, 144–159.

MacNeilage, P. F. (1985). Serial-ordering errors in speech and typing. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 193–201). New York: Academic.

Pinet, S., Ziegler, J. C., and Alario, F. X. (2016). Typing is writing: Linguistic properties modulate typing execution. *Psychonomic Bulletin and Review, 23(6),* 1898–1906. https://doi.org/10.3758/s13423-016-1044-3

Shaffer, L. H. (1975). Control processes in typing. *Quarterly Journal of Experimental Psychology, 27*, 419–432.

Tannenbaum, P. H., Williams, F., and Wood, B. S. (1967). Hesitation Phenomena and Related Encoding Characteristics in Speech and Typewriting. *Language and Speech, 10(3)*, 203–215. https://doi.org/10.1177/002383096701000306

van Nes, F. L. (1976). Analysis of keying errors. *Ergonomics, 19*, 165–174.

ADDITIONAL SOURCES:

Computer Hope. (2021, July 6). When was the first keyboard invented? *Computer Hope*. Retrieved September 7, 2022, from https://www.computerhope.com/issues/ch001802.htm

Diffusion of Personal Computing Devices, 1977-2020. (2022, January 24). *The Geography of Transport Systems | the Spatial Organization of Transportation and Mobility*. Retrieved September 7, 2022, from https://transportgeography.org/contents/chapter1/the-setting-of-global-transportation-systems/personal-computing-devices/

Gutman-Wei, R. (2019, December 12). When Did Ancient Humans Start to Speak? *The Atlantic*. Retrieved September 7, 2022, from https://www.theatlantic.com/science/archive/2019/12/when-did-ancient-humans-start-speak/603484/

Kafka, F. *Metamorphosis*. Project Gutenberg, Aug 17, 2005, https://www.gutenberg.org/ebooks/5200

Merriam-Webster. (n.d.). Keylogger. In *Merriam-Webster*. Retrieved January 8, 2023, from https://www.merriam-webster.com/dictionary/keylogger

Moody, R. (2022, March 21). Screen Time Statistics: Average Screen Time in US vs. the rest of the world. *Comparitech*. Retrieved September 6, 2022, from https://www.comparitech.com/tv-streaming/screen-time-statistics/

Appendix A

The typing test

https://www.keyhero.com/custom-typing-test/metamorphosisexcerpt/

Title: MetamorphosisExcerpt

Text: After a while he had already moved so far across that it would have been hard for him to keep his balance if he rocked too hard. The time was now ten past seven and he would have to make a final decision very soon. Then there was a ring at the door of the flat. “That’ll be someone from work”, he said to himself, and froze very still, although his little legs only became all the more lively as they danced around. For a moment everything remained quiet. “They’re not opening the door”, Gregor said to himself, caught in some nonsensical hope. But then of course, the maid’s firm steps went to the door as ever and opened it. Gregor only needed to hear the visitor’s first words of greeting and he knew who it was—the chief clerk himself. Why did Gregor have to be the only one condemned to work for a company where they immediately became highly suspicious at the slightest shortcoming?

Were all employees, every one of them, louts, was there not one of them who was faithful and devoted who would go so mad with pangs of conscience that he couldn’t get out of bed if he didn’t spend at least a couple of hours in the morning on company business? Was it really not enough to let one of the trainees make enquiries—assuming enquiries were even necessary—did the chief clerk have to come himself, and did they have to show the whole, innocent family that this was so suspicious that only the chief clerk could be trusted to have the wisdom to investigate it? And more because these thoughts had made him upset than through any proper decision, he swang himself with all his force out of the bed. There was a loud thump, but it wasn’t really a loud noise. His fall was softened a little by the carpet, and Gregor’s back was also more elastic than he had thought, which made the sound muffled and not too noticeable. He had not held his head carefully enough, though, and hit it as he fell; annoyed and in pain, he turned it and rubbed it against the carpet.

Appendix B

The survey

https://forms.gle/iv9h3UA1ZvXWyvzV6

Exit survey

What gender do you identify as?

- Male
- Female
- Prefer not to say
- Other: ______________

What is your age?

____

What country/state do you live in?

____

What is the highest degree or level of education you have completed?

- Some high school
- High school
- Bachelor's degree
- Master's degree
- Ph.D. or higher
- Trade school
- Other: ______________

How would you rate your typing skills?

	1	2	3	4	5	
Far below average	O	O	O	O	O	Far above average

Which keyboard layout do you use?

- QWERTY
- QWERTZ
- AZERTY
- DVORAK
- COLEMAK
- WORKMAN

- MTGAP
- Other: _____________

Please upload the results from the “MetamorphosisExcerpt [1/2]” typing test here.

Please upload the results from the “MetamorphosisExcerpt [2/2]” typing test here.

Appendix C

Complete participant demographic data

Table C1

Participant Demographic Data

No.	Gender	Age	Place of residence	Level of education
1	Female	34	USA	Master's degree
2	Female	22	USA	Bachelor's degree
3	Male	24	USA	Associate degree
4	Nonbinary	24	USA	High school
5	Nonbinary	17	USA	Some high school
6	Female	19	Canada	High school
7	Male	39	USA	Bachelor's degree
8	Male	19	USA	Bachelor's degree
9	Female	29	UK	Bachelor's degree
10	Male	23	Denmark	High school
11	Male	42	South Africa	Bachelor's degree
12	Female	34	South Africa	Master's degree
13	Female	41	South Africa	Master's degree
14	Male	34	South Africa	Master's degree
15	Male	17	USA	Some high school
16	Female	28	USA	Bachelor's degree
17	Male	25	Philippines	Bachelor's degree
18	Female	34	Australia	Master's degree
19	Female	30	South Africa	Bachelor's degree
20	Female	23	USA	Bachelor's degree
21	Female	30	South Africa	Bachelor's degree
22	Female	23	USA	Bachelor's degree
23	Female	30	South Africa	Bachelor's degree
24	Female	21	Canada	High school
25	Female	27	UK	Bachelor's degree

Appendix D

Error marking

Table D1

Formatting Styles

Formatting	Description	Example
italicized	This formatting marks the error unit(s) in the word.	enquirie*r*s
bolded and underlined	This formatting marks the corrected error segment(s) in the original *Key Hero* error overview.	ev<u>**ry**</u>ery<u>**one** </u>y one
bolded and crossed out	This formatting marks the uncorrected error segment(s) in the original *Key Hero* error overview.	susp**~~ocious~~**

Table D2

Additional Symbols

Symbol	Description	Example
_	Marks a missing unit	c_uple of hours
[	Marks the moment in which the typist stopped typing out the intended string in order to correct an error. The text following this symbol is the (assumed) intended string.	*b*o[mpany business]
]	Marks the end of the (assumed) intended string.	*b*o[mpany business]
· [14]	Marks an empty space in a location where it should not be.	himself·,

[14] This symbol is not a full stop but an interpunct or centered dot.

Appendix E

The uncorrected error subcorpus

Table E1

Monopositional Contextual Typing Errors

Error class	Error text	Target text
anticipatory contextual within-word substitution	b*i*siness	business
anticipatory contextual within-word substitution	trainees m*e*ke enquiries	trainees make enquiries
nonadjacent anticipatory contextual within-word masking	—_id the	—did the
nonadjacent perseveratory contextual within-word masking	becaus_ these	because these
nonadjacent perseveratory contextual within-word masking	Grego_'s	Gregor's
adjacent contextual within-word addition	too hard*d*.	too hard.
adjacent contextual within-word addition	the doo*o*r",	the door",
adjacent contextual within-word addition	really*y* a loud	really a loud
adjacent contextual within-word addition	he didn't*t*	he didn't spend
adjacent contextual within-word addition	clerk hav*v* to come	clerk have to come
adjacent contextual within-word addition	investigate*e* it	investigate it
adjacent contextual within-word addition	elastic t*t*han	elastic than
nonadjacent contextual within-word addition	remain*d*ed quiet	remained quiet
nonadjacent contextual within-word addition	m*e*ake enquiries	trainees make enquiries
perseveratory contextual between-word masking	really a _oud	really a loud
anticipatory contextual between-word masking	a f_nal decision	a final decision
anticipatory contextual between-word masking	rema_ned quiet	remained quiet
anticipatory contextual between-word masking	—the _hief clerk	—the chief clerk
anticipatory contextual between-word masking	c_uple of hours	couple of hours
contextual between-word addition	enough*t* to let	enough to let
contextual between-word addition	wasn't*r* really	wasn't really

Table E2

Monopositional Non-contextual Typing Errors

Error class	Error text	Target text
non-contextual addition	beca*u*me all the more	became all the more
non-contextual omission	a _ouple	a couple
non-contextual omission	enough to _et	enough to let
non-contextual omission	enoug_ to let	enough to let
non-contextual omission	easti_ than	elastic than
non-contextual omission	elasti_ than	elastic than
non-contextual omission	have _he	have the
non-contextual omission	investi_ate it	investigate it
non-contextual omission	all h_s force	all his force

Table E3

Bipositional Typing Errors

Error class	Error text	Target text
C1C2 → C2C1 adjacent switch	hi*sm*elf	himself
V1V2 → V2V1 adjacent switch	enquir*ei*s	enquiries
CV → VC adjacent switch	*ot* the door	to the door
CV → VC adjacent switch	only o*en* condemned	only one condemned
CV → VC adjacent switch	t*ah*t it	that it

Appendix F

The corrected error subcorpus

Table F1

Monopositional Contextual Typing Errors

Error class	Error text	Target text
anticipatory within-word substitution	a *d*[oud noise]	a loud noise
anticipatory within-word substitution	a*d*d	and
anticipatory within-word substitution	c*a*mpany	company
anticipatory within-word substitution	For a mo*n*[ent]	a moment
anticipatory within-word substitution	*l*[ore lively]	more lively
anticipatory within-word substitution	little *g*[egs]	his little legs
anticipatory within-word substitution	*q*n[quiries]	enquiries
anticipatory within-word substitution	*s*[onscience]	conscience
anticipatory within-word substitution	s*i*[spicious]	suspicious
anticipatory within-word substitution	*t*[ith]	with
anticipatory within-word substitution	The*r*'re not	They're not
anticipatory within-word substitution	too n*i*tic[eable]	too noticeable
anticipatory within-word substitution	too not*e*[ceable]	too noticeable
anticipatory within-word substitution	too not*e*[ceable]	not too noticeable
anticipatory within-word substitution	too not*e*[ceable]	too noticeable
perseveratory within-word substitution	cond*o*[mned]	condemned
perseveratory within-word substitution	final deci*c*ion	final decision
perseveratory within-word substitution	it w*w*s	it was
perseveratory within-word substitution	open*e*n[g]	opening
perseveratory within-word substitution	suspicio*i*s	suspicious
perseveratory within-word substitution	they immedi*e*[tely]	they immediately
perseveratory within-word substitution	too notic*o*[able]	too noticeable
adjacent within-word masking	al_ e[mployees]	all employees
adjacent within-word masking	an_oyed	annoyed
adjacent within-word masking	—as_uming	—assuming
adjacent within-word masking	employe_s	employees
adjacent within-word masking	neces_a[ry]	necessary
adjacent within-word masking	neces_a[ry]	necessary
adjacent within-word masking	not to_ no[ticeable]	not too noticeable
adjacent within-word masking	the traine_s	the trainees
adjacent within-word masking	to_[noticeable]	not too noticeable

adjacent within-word masking	Was it real_y	Was it really
nonadjacent anticipatory within-word masking	a _oment	a moment
nonadjacent anticipatory within-word masking	final dec_sion	final decision
nonadjacent anticipatory within-word masking	s_me[one]	someone
nonadjacent anticipatory within-word masking	susp_c[ious]	suspicious
nonadjacent perseveratory within-word masking	and di_	and did
nonadjacent perseveratory within-word masking	condem_ed	condemned
nonadjacent perseveratory within-word masking	condem_ed	condemned
nonadjacent perseveratory within-word masking	condemn_d	condemned
nonadjacent perseveratory within-word masking	cons_i[ence]	conscience
nonadjacent perseveratory within-word masking	cons_i[ence]	conscience
nonadjacent perseveratory within-word masking	cons_i[ence]	conscience
nonadjacent perseveratory within-word masking	enquir_es	enquiries
nonadjacent perseveratory within-word masking	enquir_es	make enquiries
nonadjacent perseveratory within-word masking	enquir_es	enquiries
nonadjacent perseveratory within-word masking	ev_ry[one]	every one
nonadjacent perseveratory within-word masking	invest_g[ate it]	investigate it
nonadjacent perseveratory within-word masking	make enquir_es	make enquiries
nonadjacent perseveratory within-word masking	neede_ to	needed to
nonadjacent perseveratory within-word masking	only becam_	only became
nonadjacent perseveratory within-word masking	only neede_ to hear	only needed
nonadjacent perseveratory within-word masking	remain_d quiet	remained quiet
nonadjacent perseveratory within-word masking	soften_d	softened
nonadjacent perseveratory within-word masking	soften_d	softened
nonadjacent perseveratory within-word masking	step_[went]	steps went
nonadjacent perseveratory within-word masking	su_p[icious]	suspicious
nonadjacent perseveratory within-word masking	su_pic[ious]	suspicious
nonadjacent perseveratory within-word masking	su_picious	suspicious
nonadjacent perseveratory within-word masking	suspic_ous	suspicious
nonadjacent perseveratory within-word masking	suspicio_s	suspicious
nonadjacent perseveratory within-word masking	the vis_t[or's]	the visitor's
nonadjacent perseveratory within-word masking	they immediat_ly	they immediately
nonadjacent perseveratory within-word masking	thoug_t	thought
nonadjacent perseveratory within-word masking	though_,	thought,
nonadjacent perseveratory within-word masking	though_, wh[ich]	thought, which
nonadjacent perseveratory within-word masking	though_, wh[ich]	thought, which
adjacent within-word addition	a little*e*	a little
adjacent within-word addition	a*a*[ring at]	a ring at

adjacent within-word addition	a*a*nd	and
adjacent within-word addition	agains*s*t	against
adjacent within-word addition	an*n*[d]	and
adjacent within-word addition	any*y* [proper]	any proper
adjacent within-word addition	—ass*s*[uming]	—assuming
adjacent within-word addition	at lea*a*st a couple	at least
adjacent within-word addition	beca*a*[use]	because
adjacent within-word addition	decisio*o*[n]	decision
adjacent within-word addition	de*e*[vote]	devoted
adjacent within-word addition	enqui*i*[ries]	enquiries
adjacent within-word addition	enquiries*s*	enquiries
adjacent within-word addition	*g*G[regor]	Gregor
adjacent within-word addition	go s*s*o [mad]	go so mad
adjacent within-word addition	gree*e*[ting]	greeting
adjacent within-word addition	have the*e*	have the
adjacent within-word addition	have*e*[to]	have to
adjacent within-word addition	h*h*e [swang]	he swang
adjacent within-word addition	hi*i*m[self]	himself
adjacent within-word addition	hi*i*m[self]	himself
adjacent within-word addition	hi*i*mse[lf]	himself
adjacent within-word addition	innocent f*f*[amily]	innocent family
adjacent within-word addition	make enqui*i*[ries]	enquiries
adjacent within-word addition	not enoug*g*h	not enough
adjacent within-word addition	noticea*a*[ble]	noticeable
adjacent within-word addition	of cour*r*se	of course
adjacent within-word addition	proper*r*	proper
adjacent within-word addition	really*y*	Was it really
adjacent within-word addition	show*w*[the whole]	show the whole
adjacent within-word addition	susp*p*[icious]	suspicious
adjacent within-word addition	sus*s*[picious]	suspicious
adjacent within-word addition	sus*s*[picious]	suspicious
adjacent within-word addition	swang*g*	swang
adjacent within-word addition	swang*g* [himself]	swang himself
adjacent within-word addition	swang*g* him[self]	swang himself
adjacent within-word addition	swang*g*[himself]	swang himself
adjacent within-word addition	the carpet*t*[, and]	the carpet, and
adjacent within-word addition	the chi*i*[ef]	the chief
adjacent within-word addition	the fla*a*t	the flat

adjacent within-word addition	the trainees*s*	the trainees
adjacent within-word addition	the t*t*ra[inees]	trainees
adjacent within-word addition	the visitor's*s*	the visitor's first
adjacent within-word addition	to make*e*[a final]	to make a final
adjacent within-word addition	to show*w*	to show
adjacent within-word addition	too*o*[hard]	too hard
adjacent within-word addition	t*t*[oo]	too
adjacent within-word addition	very s*s*oo[n]	very soon
adjacent within-word addition	was sof*f*[tened]	was softened
adjacent within-word addition	wasn't*t* [really]	wasn't really
adjacent within-word addition	went*t* [to]	went to
adjacent within-word addition	Were all*l*	Were all employees
adjacent within-word addition	who wo*o*uld	who would
adjacent within-word addition	W*W*[hy did]	Why did
nonadjacent within-word addition	condem*e*ned	condemned
nonadjacent within-word addition	enough, though*t*	enough, though
nonadjacent within-word addition	enq*i*uir[ies]	enquiries
nonadjacent within-word addition	enquir*e*ie[s]	enquiries
nonadjacent within-word addition	G*e*[regor]	Gregor
nonadjacent within-word addition	G*e*[regor's]	Gregor's
nonadjacent within-word addition	hel*e*d h[is head]	held his head
nonadjacent within-word addition	his *a*balanc[e]	his balance
nonadjacent within-word addition	his balan*a*[ce]	his balance
nonadjacent within-word addition	little legs*l*	little legs
nonadjacent within-word addition	make enqui*e*[ries]	enquiries
nonadjacent within-word addition	m*e*ore [because]	more because
nonadjacent within-word addition	mov*o*ed so	moved so
nonadjacent within-word addition	opene*n*d[it]	opened it
nonadjacent within-word addition	proper decisi*s*on	proper decision
nonadjacent within-word addition	prop*o*er	proper
nonadjacent within-word addition	short*o*[coming]	shortcoming
nonadjacent within-word addition	slight*l*[est]	slightest
nonadjacent within-word addition	slight*l*[est]	slightest
nonadjacent within-word addition	step*e*s[went]	steps went
nonadjacent within-word addition	sus*c*[picious]	suspicious
nonadjacent within-word addition	suspi*s*c[ious]	suspicious
nonadjacent within-word addition	through*t*[any]	through any
nonadjacent within-word addition	*v*eve[ry one]	every one

nonadjacent within-word addition	which*i*[made]	which made
anticipatory between-word substitution	and *t*[id they]	did they
anticipatory between-word substitution	*b*o[mpany business]	company business
anticipatory between-word substitution	eve*n*[and]	ever and
anticipatory between-word substitution	hit i*s* [as]	hit it as
anticipatory between-word substitution	h*o* [couldn't]	he couldn't
anticipatory between-word substitution	kn*o*w [who]	knew how
anticipatory between-word substitution	*n*[eally not]	really not
anticipatory between-word substitution	no*t*[ten]	now ten
anticipatory between-word substitution	no*t*[ten]	now ten
anticipatory between-word substitution	o*g* [greeting]	of greeting
anticipatory between-word substitution	*s*[ery still]	very still
anticipatory between-word substitution	t[id they]	did they
anticipatory between-word substitution	to *t*[how the]	to show the
anticipatory between-word substitution	tursted *h*[o have]	trusted to have
anticipatory between-word substitution	wh*a*[ch made]	which made
perseveratory between-word substitution	greeting and he *g*n[ew]	greeting and he knew
perseveratory between-word substitution	Gregor *e*[only]	Gregor only
perseveratory between-word substitution	he roc*h*[ed] too hard	he rocked too hard
perseveratory between-word substitution	him to *h*e[ep his]	him to keep his
perseveratory between-word substitution	let on*t*	let one
perseveratory between-word substitution	let on*t* [of]	let one of
perseveratory between-word substitution	of be*f* [if]	of bed if
perseveratory between-word substitution	pangs of *s*[onscience]	pangs of conscience
perseveratory between-word substitution	slightest shortcom*e*[ng]	slightest shortcoming
perseveratory between-word substitution	so s*o*[spicious]	so suspicious
perseveratory between-word substitution	some *s*o[nsensical]	some nonsensical
perseveratory between-word substitution	that th*a*[s was]	this was
perseveratory between-word substitution	the *t*[oor]	the door
perseveratory between-word substitution	thoughts ha*s*	thoughts had
perseveratory between-word substitution	thoughts ha*s*	thoughts had
perseveratory between-word substitution	to hims*o*lf[,]	to himself,
perseveratory between-word substitution	to h*o*mself,	to himself,
perseveratory between-word substitution	upset tha*t* [through]	upset than through
perseveratory between-word substitution	visitor's first *v*[ords]	visitor's first words

anticipatory between-word masking	hi_ it	hit it
anticipatory between-word masking	the _hief clerk	the chief clerk
anticipatory between-word masking	wa_[softened]	was softened
perseveratory between-word masking	all emp_oyees	all employees
perseveratory between-word masking	and _evoted	and devoted
perseveratory between-word masking	door _f	door of
perseveratory between-word masking	greeting a_d	greeting and
perseveratory between-word masking	he swang _i[mself]	swang himself
perseveratory between-word masking	knew _ho	knew who
perseveratory between-word masking	lively as the_	lively as they
perseveratory between-word masking	slightest _h[ortcoming]	slightest shortcoming
perseveratory between-word masking	slightest shor_coming	slightest shortcoming
perseveratory between-word masking	so _us[picious]	suspicious
perseveratory between-word masking	that _his[was]	that this was
perseveratory between-word masking	the c_ief	the chief
perseveratory between-word masking	the c_ief	the chief
perseveratory between-word masking	the c_ief clerk	the chief clerk
perseveratory between-word masking	these t_o[ughts had]	these thoughts had
perseveratory between-word masking	thoughts _a[d made]	thoughts had made
perseveratory between-word masking	Was it re_l[ly]	Was it really
perseveratory between-word masking	who w_ul[d]	who would
plain contextual between-word omission	_i[t it]	hit it
plain contextual between-word omission	_i[t it]	hit it
plain contextual between-word omission	at lea_t	at least
plain contextual between-word omission	Then ther_	Then there
between-word addition	a couple of hours*e*	a couple of hours
between-word addition	all the *t*[more]	the more lively
between-word addition	couple of hours*e*	couple of hours
between-word addition	enough t*h*[o]	enough to
between-word addition	enough*t* to	enough to
between-word addition	enough*t*, [though]	enough, though
between-word addition	enough*t*[, though]	enough, though
between-word addition	have *e*the	have the
between-word addition	have the wisdom*e*	have the wisdom
between-word addition	have the wisdom*e*	have the wisdom

between-word addition	have the wisdom*e*	have the wisdom
between-word addition	have*t* to	have to
between-word addition	hit *h*[it as he]	hit it as he
between-word addition	hit it *h*a[s he]	hit it as he
between-word addition	little *e*b[y]	little by
between-word addition	maid's firm ste*m*[ps]	maid's firm steps
between-word addition	rubbe*n*[d]	and rubbed
between-word addition	said *o*[to himself]	said to himself
between-word addition	show the *s*w[hole]	show the whole
between-word addition	*t*[he turned]	he turned
between-word addition	the wisdom*e* [to]	the wisdom to
between-word addition	the wis*o*dom	the wisdom
between-word addition	there*w*[was]	there was
between-word addition	through *t*[any]	through any
between-word addition	*t*of [them]	of them
between-word addition	trainees make*s* en[quiries]	trainees make enquiries
between-word addition	trainees make*s*[enquiries]	trainees make enquiries
between-word addition	wisdom *w*[to]	wisdom to

Table F2

Monopositional Noncontextual Typing Errors

Error class	Error text	Target text
noncontextual substitution	bed i*s* h[e didn't spend]	bed if he
noncontextual substitution	hel*p*	held
noncontextual substitution	mad with pan*d*s	with pangs
noncontextual substitution	mor*e*[ning]	morning
noncontextual substitution	the doo*s*	the door
noncontextual substitution	there wa*y* a	there was a
noncontextual substitution	wisdom t*h*	wisdom to
noncontextual omission	_a[ke enquiries]	make enquiries
noncontextual omission	_a[sn't really]	wasn't really
noncontextual omission	_a[ve to]	have to
noncontextual omission	_as	was
noncontextual omission	_as sof[tened]	was softened
noncontextual omission	_e[ally]	really

noncontextual omission	_h[e chief]	the chief clerk
noncontextual omission	_h[e flat]	the flat
noncontextual omission	_h[e maid's firm steps]	the maid's firm steps
noncontextual omission	_w[ang himself]	swang himself
noncontextual omission	a _ou[ple]	a couple
noncontextual omission	a cou_l[e]	a couple
noncontextual omission	a fi_a[l decision]	a final decision
noncontextual omission	a fi_al [decision]	a final decision
noncontextual omission	a final de_i[sion]	a final decision
noncontextual omission	a_d	and
noncontextual omission	a_d	and
noncontextual omission	And _o[re]	And more
noncontextual omission	And mor_	And more
noncontextual omission	be tr_s[ted]	trusted
noncontextual omission	be_[if he]	bed if he
noncontextual omission	but _t	but it
noncontextual omission	carefull_[enough]	carefully enough
noncontextual omission	clerk _ave	clerk have
noncontextual omission	clerk _i[mself]	clerk himself
noncontextual omission	clerk _imself	clerk himself
noncontextual omission	com_	come
noncontextual omission	come _imsel[f]	come himself
noncontextual omission	company bus_ne[ss]	company business
noncontextual omission	could_'[t]	couldn't
noncontextual omission	eve_y	every
noncontextual omission	eve_ything	everything
noncontextual omission	ever_thing	everything remained
noncontextual omission	f_rm ste[ps]	firm steps
noncontextual omission	fell; _n[noyed]	fell; annoyed
noncontextual omission	for_e	force
noncontextual omission	h_v[e]	have
noncontextual omission	hav_ to	have to
noncontextual omission	he _wa[ng]	he swang
noncontextual omission	he s_a[ng]	he swang
noncontextual omission	he sa_d	he said
noncontextual omission	he sa_d to	he said to
noncontextual omission	he sai_	he said
noncontextual omission	he sw_ng	he swang

noncontextual omission	his balanc_ if	his balance if
noncontextual omission	hour_s	hours
noncontextual omission	in_e[stigate it]	to investigate it
noncontextual omission	it wa_n['t]	it wasn't
noncontextual omission	keep _is	keep his
noncontextual omission	l_ut[s, was there]	louts, was there
noncontextual omission	little le_s	little legs
noncontextual omission	m_r[e]	more
noncontextual omission	made him _p[set]	made him upset
noncontextual omission	mor_[ing]	morning
noncontextual omission	mor_[ing]	morning
noncontextual omission	more beca_se	more because
noncontextual omission	more li_e[ly]	more lively
noncontextual omission	no_ ten	now ten
noncontextual omission	not eno_gh t[o]	not enough to
noncontextual omission	not enou_h	not enough
noncontextual omission	notic_a[ble]	noticeable
noncontextual omission	on_ of the	one of the
noncontextual omission	on_y	only
noncontextual omission	one _f[them]	one of them
noncontextual omission	one of _h[em]	one of them
noncontextual omission	one of t_e	let one of the
noncontextual omission	pan_s of	pangs of
noncontextual omission	pangs of con_c[ience]	pangs of conscience
noncontextual omission	proper de_i[sion]	proper decision
noncontextual omission	proper deci_[sion]	proper decision
noncontextual omission	qu_e[t]	quiet
noncontextual omission	qu_et	quiet
noncontextual omission	qu_et	quiet
noncontextual omission	remaine_[quiet]	remained quiet
noncontextual omission	said to _im[self]	said to himself
noncontextual omission	sh_w	show
noncontextual omission	show _he	show the
noncontextual omission	slightest sho_t[coming]	slightest shortcoming
noncontextual omission	slightest shor_coming	slightest shortcoming
noncontextual omission	steps w_nt	steps went
noncontextual omission	still, _lth[ough]	still, although
noncontextual omission	t_a[t it would have]	that it would have

noncontextual omission	T_e[re was]	There was
noncontextual omission	ten p_st[seven]	ten past seven
noncontextual omission	th_o[ugh]	through
noncontextual omission	That'll _e	That'll be
noncontextual omission	the ca_p[et]	the carpet
noncontextual omission	the sou_d [muffled]	the sound muffled
noncontextual omission	the wi_d[om]	the wisdom
noncontextual omission	the wis_om	the wisdom
noncontextual omission	the_ [of course]	then of course
noncontextual omission	The_ t[here]	Then there
noncontextual omission	The_'[re not]	They're not
noncontextual omission	thump, _u[t it]	thump, but it
noncontextual omission	thump, _u[t]	thump, but
noncontextual omission	time w_s	time was
noncontextual omission	to hav_	to have
noncontextual omission	to investi_a[te it]	to investigate it
noncontextual omission	too _o[ticable]	too noticeable
noncontextual omission	very s_i[ll]	very still
noncontextual omission	w_s [now]	was now
noncontextual omission	wa_[now]	was now
noncontextual omission	was fait_fu[l]	was faithful
noncontextual omission	Was it _ea[lly]	Was it really
noncontextual omission	went t_	went to
noncontextual omission	where the_	where they
noncontextual omission	who wou_d	who would
noncontextual omission	who_e, i[nnocent]	whole, innocent
noncontextual omission	words _f	words of
noncontextual omission	woul_ have	would have
noncontextual addition	a lou*n*[d thump]	a loud thump
noncontextual addition	a while *a*[he]	a while he
noncontextual addition	an*d*y[proper]	any proper
noncontextual addition	be the *w*[only]	be the only
noncontextual addition	clerk*y*[himself]	clerk himself
noncontextual addition	have the wi*r*sdom	the wisdom
noncontextual addition	his *w*[force]	his force
noncontextual addition	legs *w*[only]	legs only
noncontextual addition	sho*u*w [the whole]	to show the whole
noncontextual addition	that *r*[it]	that it

noncontextual addition	the maid*n*['s firm steps]	maid's firm steps
noncontextual addition	th*e*at [it]	that it
noncontextual addition	to *t*[be the]	to be the
noncontextual addition	wa[s he fell]	as he fell
noncontextual addition	Were*n*[all]	Were all
noncontextual addition	w*g*ou[ld have]	would have
noncontextual addition	Why*v*[did]	Why did

Table F3

Bipositional Typing Errors

Error class	Error text	Target text
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	a ri*gn*[at]	a ring at
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	against *ht*e c[arpet]	against the carpet
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	agai*sn*t[the carpet]	against
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	be *rt*usted	trusted
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	cau*hg*[t]	caught
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	devoted *hw*o[would]	devoted who would
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	hi*hg*l[y]	highly
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	himse*fl*.	himself.
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	hw[ole]	whole
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	*hw*ere	company where
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	of *ht*e	of the
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	on*yl* [one]	only one
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	opening *ht*e d[oor]	opening the door
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	real*yl*	really
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	that on*yl* t[he]	that only the
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	the wi*d*[*s*om to]	the wisdom to
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	thou*h[g]*	though
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	thou*h[g]*	though
$C_1C_2 \rightarrow C_2C_1$ adjacent switch	whi*hc*	which
$V_1V_2 \rightarrow V_2V_1$ adjacent switch	bec*ua*se	because
$V_1V_2 \rightarrow V_2V_1$ adjacent switch	the ch*ei*[f clerk]	chief clerk
CV $\rightarrow$ VC adjacent switch	Af*et*r a	after a
CV $\rightarrow$ VC adjacent switch	bu*is*[ness]	business
CV $\rightarrow$ VC adjacent switch	danced a*or*und	around
CV $\rightarrow$ VC adjacent switch	even *en*c[essary]	even necessary

CV → VC adjacent switch	everything *er*[mained]	everything remained
CV → VC adjacent switch	everything re*am*in[ed]	remained
CV → VC adjacent switch	far ac*or*ss	far across
CV → VC adjacent switch	final *ed*[cision]	final decision
CV → VC adjacent switch	f*or*ze	froze
CV → VC adjacent switch	f*or*ze	froze
CV → VC adjacent switch	he *as*[id]	he said
CV → VC adjacent switch	investi*ag*te	investigate
CV → VC adjacent switch	ope*in*[ng the door]	opening the door
CV → VC adjacent switch	pangs of *oc*[nscience]	pangs of conscience
CV → VC adjacent switch	said *ot*	said
CV → VC adjacent switch	t*eh*[door]	the door
CV → VC adjacent switch	That'll *eb*	That'll be
CV → VC adjacent switch	the *od*or	the door
CV → VC adjacent switch	the wis*od*[m to]	the wisdom to
CV → VC adjacent switch	to *eh*ar	to hear
CV → VC adjacent switch	to inves*it*gate	to investigate
CV → VC adjacent switch	too no*it*[ceable]	too noticeable
CV → VC adjacent switch	*am*[ke enquiries]	make enquiries
VC → CV adjacent switch	a comp*na*y	a company
VC → CV adjacent switch	a m*mo*[ent]	a moment
VC → CV adjacent switch	be tr*su*[ted]	trusted
VC → CV adjacent switch	beca*su*[e]	because
VC → CV adjacent switch	condemn*de*	condemned
VC → CV adjacent switch	condemn*de*	condemned
VC → CV adjacent switch	consci*ne*[ce]	conscience
VC → CV adjacent switch	d*di*	did
VC → CV adjacent switch	didn't sp*ne*[d]	didn't spend
VC → CV adjacent switch	d*na*ced	danced
VC → CV adjacent switch	el*sa*[tic]	elastic
VC → CV adjacent switch	*fo* th[em]	of them
VC → CV adjacent switch	F*ro*[a moment]	For a
VC → CV adjacent switch	him ups*te*	him upset
VC → CV adjacent switch	hit *ti*	hit it
VC → CV adjacent switch	innocent f*ma*i[ly]	innocent family
VC → CV adjacent switch	innoc*ne*t	innocent
VC → CV adjacent switch	maid's f*ri*m [steps]	maid's firm steps
VC → CV adjacent switch	op*ne*[ed it]	opened it

VC → CV adjacent switch	that *ti*[would]	that it would
VC → CV adjacent switch	the chi*fe*	the chief
VC → CV adjacent switch	to he*ra*	to hear
VC → CV adjacent switch	was *la*so	was also
VC → CV adjacent switch	who *ti* was	who it was
misordered adjacent grapheme doubling	a lit*ll*[e]	a little
misordered adjacent grapheme doubling	all employ*ess*	all employees
misordered adjacent grapheme doubling	all employ*ess*	all employees
misordered adjacent grapheme doubling	—a*suu*ming	—assuming
misordered adjacent grapheme doubling	employ*ess*	employees
misordered adjacent grapheme doubling	even nece*saa*[ry]	even necessary
misordered adjacent grapheme doubling	gr*ett*i[ng]	greeting
misordered adjacent grapheme doubling	gr*ett*in[g]	greeting
misordered adjacent grapheme doubling	His f*aal*	His fall
misordered adjacent grapheme doubling	*sso*n	soon
misordered adjacent grapheme doubling	the d*orr*	the door
single grapheme reversal	a mom*t*n*[e]*	a moment
single grapheme reversal	were e*n[ev]*	were even
anticipatory shift	more li*l*ve[y]	more lively
anticipatory shift	proper de*i*ci[son]	proper decision
anticipatory shift	proper de*i*ci[son]	proper decision
anticipatory shift	the w*d*iso[m to]	the wisdom to
perseveratory shift	concien*s*[ce]	conscience
perseveratory shift	suspciou*i*[s]	suspicious