Observing and analysing digital writing processes with Inputlog

Mariëlle Leijten & Luuk Van Waes

Introduction

- Faculty of Applied Economics
- Department of Management
- Teaching: Business Communication (Dutch)
- Research group on ‘Writing and Digital Media’
  - writing from multiple (digital) sources
  - reading during writing
  - speech recognition & live subtitling
  - online Writing Center (Calliope)

Introduction

- overview
- current version of Inputlog (+ demo)
- research studies
  - typing errors ► micro
  - live subtitling ► meso
  - multiple sources ► macro
- current developments

Anecdote

Keystroke logging: general
**Objective**

Study writing as it unfolds in real time:
- registration device
- representation of the registration
- framework and procedures for the analysis

**Methods in writing research**

- product analysis
- thinking aloud protocols: simultaneous or retrospective
- second or triple task techniques: direct or indirect
- video recording
- eye tracking
- neuro imaging (fMRI – PET – ERP)
- keystroke logging

---

**KSL programs**

- Scriptlog
- Inputlog
- Translog
- uLog
- TraceIt/JEdit
- Eyewrite
- Eye&Pen

**Translog**

- especially designed for translation studies
- two windows:
  1. source text (limited number of lines)
  2. target text (Translog window)
- analyses
  - linear log file
  - dynamic replay
- further developments:
  - integration with eyetracking
  - interactive dictionaries (reading activation)

**ScriptLog**

- A never-ending story that started in 1985
- New Windows version 1.8.36
- Mainly for quasi-experimental/quasi-natural experiments
- A simple editor that records all keyboard events and mouse events and their temporal distribution.
- Could be used for text production or more controlled experiments of the production of smaller units, e.g. spelling tests etc.

http://www.scriptlog.net

**History**

- Excel Macro
- HyperCard Stack
- Mac Program + Perl & Awk scripts
- Windows program with analysis tools included
- Eye tracking
**Needs**
- Better Synchronization with the eye tracker
- More experimental possibilities – dual/triple task, more timed elicitation etc.
- A simple more robust recorder for classroom usage with playback function
- Annotation tools
- A more advanced recorder with advanced analysis tools - generating, that can import data from the simple tool
- Exportable data

**Future ScriptLog**

**Current version of Inputlog**

**Inputlog 5 beta**
- Windows (e.g. MS Word)
- Writing modes
  - keyboard and mouse movements & clicks
  - speech: Dragon Naturally Speaking
  - focus: window monitoring
- Analyses
  - process- & product characteristics
  - linear text development and S-Notation
  - pausing behavior
  - play-back*
  - revision analysis*

**Record**
- Event selection
- Plugin
- Document selection
- Session identification

**Analyze**
- Selection of file and destination
- Configure filters
- Analyses
Summerschool Writing Process Research 2011: Keystroke logging & Eye tracking
Wednesday September 7 - Introduction to keystroke logging

Play
- Single logging session
- Session based on previous document

Play settings
- Start - stop
- Realtime
- Stepwise forward/backward
- To end/beginning of document
- In percentage of realtime speed

Convert
- Idf files made in Inputlog 4 Beta (website) can be converted to latest version of Inputlog
- Idfx files is created
- Analyses can be done:
  - General
  - Linear
  - Pause
  - Summary

Remark:
- to use Revision analyses logging sessions need to be recorded with Inputlog 5.*

General logging file

<table>
<thead>
<tr>
<th>Event</th>
<th>Output</th>
<th>Position</th>
<th>Duration</th>
<th>StartTime</th>
<th>EndTime</th>
</tr>
</thead>
<tbody>
<tr>
<td>focus</td>
<td>TASKBAR</td>
<td></td>
<td></td>
<td>6781</td>
<td>0:00:06</td>
</tr>
<tr>
<td>focus</td>
<td>Microsoft PowerPoint -</td>
<td></td>
<td></td>
<td>8409</td>
<td>0:00:08</td>
</tr>
<tr>
<td>mouse</td>
<td>Left Click</td>
<td></td>
<td></td>
<td>9391</td>
<td>0:00:09</td>
</tr>
<tr>
<td>focus</td>
<td>WordLog.docx</td>
<td>1</td>
<td>0</td>
<td>11259</td>
<td>0:00:11</td>
</tr>
<tr>
<td>keyboard</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>11259</td>
<td>0:00:11</td>
</tr>
<tr>
<td>keyboard</td>
<td>e</td>
<td>1</td>
<td>1</td>
<td>11259</td>
<td>0:00:12</td>
</tr>
<tr>
<td>keyboard</td>
<td>s</td>
<td>2</td>
<td>2</td>
<td>13281</td>
<td>0:00:13</td>
</tr>
<tr>
<td>keyboard</td>
<td>G</td>
<td>3</td>
<td>3</td>
<td>14075</td>
<td>0:00:14</td>
</tr>
<tr>
<td>keyboard</td>
<td>BACK</td>
<td>4</td>
<td>4</td>
<td>15000</td>
<td>0:00:15</td>
</tr>
<tr>
<td>keyboard</td>
<td>t</td>
<td>3</td>
<td>2</td>
<td>15891</td>
<td>0:00:15</td>
</tr>
<tr>
<td>focus</td>
<td>TASKBAR</td>
<td>16041</td>
<td>0:00:16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>focus</td>
<td>PowerPoint -</td>
<td></td>
<td></td>
<td>16020</td>
<td>0:00:16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Leijen/variables.ppt</td>
<td></td>
</tr>
<tr>
<td>mouse</td>
<td>Left Click</td>
<td>18932</td>
<td>0:00:18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>focus</td>
<td>TASKBAR</td>
<td>32196</td>
<td>0:00:31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Linear text

S-Notation
While composing a text, writers often reread their emerging text in order to complete what they intended to say, but when they detect an error they sometimes prioritize error correction.

REVISIONS
1: Deletion [...]
2: Insertion {...}

Case Typing errors
...
Typing errors


Donec commodo, est vitae eulknif rutrum, magna quam faucibus, et eget pellentesque libero odio id nunc. Pellentesque nunc nisl, congue vitae dignissim et, aliquam eu augue. Nam vel odio hendrerit elit vehicula bibendum id non mauris. Aenean urna lorem, pulvinar at orci nec, aelit id scetius odio. Morbi nec sem eros, tincidunt eu lobortis accumsan. Sed vehicula fermentum lectus, sit amet lacinia diam placerat sit amet.


Typing errors


Donec commodo, est vitae eulknif rutrum, magna quam faucibus, et eget pellentesque libero odio id nunc. Pellentesque nunc nisl, congue vitae dignissim et, aliquam eu augue. Nam vel odio hendrerit elit vehicula bibendum id non mauris. Aenean urna lorem, pulvinar at orci nec, aelit id scetius odio. Morbi nec sem eros, tincidunt eu lobortis accumsan. Sed vehicula fermentum lectus, sit amet lacinia diam placerat sit amet.


Research objectives

Theoretical
- identify online cognitive traces of typing errors
- investigate the effect of typing skills
- investigate the effect of digraph characteristics

Applied
- facilitate the identification of typing errors when conducting automatic revision analyses (Inputlog)

Method

participants 30 (15 touch typists + 15 hunt&peck)
task copy tasks (word level)
focus production of typing error
digraph frequency (CELEX)
location in word (b/e/s)
finger combination

element respective grondslagen

analysis hierarchical multilevel analysis

Materials

<table>
<thead>
<tr>
<th>digraphs</th>
<th>typing errors</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1: all</td>
<td>87614</td>
<td>2145</td>
</tr>
<tr>
<td>Set 2: only characters (×3)</td>
<td>69842</td>
<td>1535</td>
</tr>
<tr>
<td>Set 3: occurrence &gt; 500</td>
<td>50232</td>
<td>1252</td>
</tr>
<tr>
<td>Set 4: occurrence &gt; 1000</td>
<td>37747</td>
<td>681</td>
</tr>
</tbody>
</table>

Final set: 20 different digraphs
- freq. in Celex10 digr. in +25%
- adjacent: +adj 5 vs. -adj 15
- left-right: LL 5 UK 5 RR 3 RL 7

Pausing time
Data analysis

**Multilevel analysis**: hierarchical cross-classified model

![Diagram of multilevel analysis]

<table>
<thead>
<tr>
<th></th>
<th>Est.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>162</td>
<td>7.7</td>
</tr>
<tr>
<td>typing error</td>
<td>65</td>
<td>4.9</td>
</tr>
<tr>
<td>adjacent</td>
<td>-102</td>
<td>7.3</td>
</tr>
<tr>
<td>same hand</td>
<td>57</td>
<td>16.0</td>
</tr>
<tr>
<td>imm. correction</td>
<td>41</td>
<td>9.8</td>
</tr>
</tbody>
</table>

Pausing time model

**Netto model**

![Table of pausing time model]

Pausing time before and after

Immediate revision:

![Diagram of immediate revision]

Delayed revision:

![Diagram of delayed revision]

Case Live Subtitling & Speech Recognition

![Diagram of case live subtitling and speech recognition]
Temporal representation of live subtitling process

**Method**

- **participants**: 12 subtitlers (8 male, 4 female)
- **Flemish Public Television (VRT)**
- **task**: live-subtitling in three reduction modes
  - - verbatim (100%)
  - - summarized (50%)
  - - highly-reduced (25%)

**Sources**

- Luyckx et al. 2010
- Source: Luyckx et al. 2010
- Source: Luyckx et al. 2010

**Example**

<table>
<thead>
<tr>
<th>Spoken Comment</th>
<th>Transcribed text</th>
<th>Subtitle</th>
</tr>
</thead>
<tbody>
<tr>
<td>ik ben al vrij, vast weken met gaan plan omdat ik geen fazie bin</td>
<td>ik ben al vrij, vast weken met gaan plan omdat ik geen fazie bin</td>
<td>2010-01-02 10:00:00</td>
</tr>
<tr>
<td>omdat ik geen fase bin vandaag</td>
<td>omdat ik geen fase bin vandaag</td>
<td>2010-01-02 10:00:01</td>
</tr>
<tr>
<td>ik ben al vrij</td>
<td>ik ben al vrij</td>
<td>2010-01-02 10:00:02</td>
</tr>
</tbody>
</table>

---

Example paraphrase/reduction error correction/delay

---

**General file**

<table>
<thead>
<tr>
<th>Case</th>
<th>Writing from multiple (digital) sources</th>
<th>2020.03.01 10:00:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>マルウララ</td>
<td>2020.03.01 10:00:00</td>
</tr>
<tr>
<td>2</td>
<td>マルウララ</td>
<td>2020.03.01 10:00:00</td>
</tr>
<tr>
<td>3</td>
<td>マルウララ</td>
<td>2020.03.01 10:00:00</td>
</tr>
<tr>
<td>4</td>
<td>マルウララ</td>
<td>2020.03.01 10:00:00</td>
</tr>
<tr>
<td>5</td>
<td>マルウララ</td>
<td>2020.03.01 10:00:00</td>
</tr>
<tr>
<td>6</td>
<td>マルウララ</td>
<td>2020.03.01 10:00:00</td>
</tr>
</tbody>
</table>

---

**Cases Writing from multiple (digital) sources**

- New simple savvy legendary bright
- Introspective social nourishing vast
- Energizing uplifting funny shared
- Beautiful provocative insightful mad
- Current dynamic engaging lasting
- Us elevating poignant bold
- Original detailed vivid giving

---

Mariëlle Leijten
Flanders Research Foundation (FWO)
University of Antwerp – marielle.leijten@ua.ac.be

Luuk Van Waes
University of Antwerp
Luuk.vanwaes@ua.ac.be
Twitter

Writing processes change.
Writing processes are becoming more concise.
Writing processes are often based on multiple sources.

Twitter is a social networking and microblogging service, enabling its users to send and read messages called tweets. Tweets are text-based posts up to 140 characters (often based on multiple digital sources).

Method

participants: master students MPC (8 male, 61 female)
task: write a tweet (max. 140 characters)
focus: process of writing multiple digital sources
duration: max. 10 minutes
analysis: visualisation sources progress analysis network analysis (pajek)

General results

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>minimum</th>
<th>maximum</th>
<th>st. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>22.2</td>
<td>20.0</td>
<td>32.0</td>
<td>1.5</td>
</tr>
<tr>
<td>Number of characters</td>
<td>408.7</td>
<td>139.0</td>
<td>861.0</td>
<td>161.5</td>
</tr>
<tr>
<td>Number of characters</td>
<td>413.8</td>
<td>155.0</td>
<td>993.0</td>
<td>179.0</td>
</tr>
<tr>
<td>Number of words</td>
<td>41.8</td>
<td>14.0</td>
<td>97.0</td>
<td>19.2</td>
</tr>
<tr>
<td>Total process time (seconds)</td>
<td>474.4</td>
<td>219.8</td>
<td>718.1</td>
<td>107.0</td>
</tr>
<tr>
<td>Total pause time (seconds)</td>
<td>92.9</td>
<td>13.4</td>
<td>305.2</td>
<td>60.2</td>
</tr>
<tr>
<td>Percentage pause time</td>
<td>19.6</td>
<td>6.9</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Ratio produced characters</td>
<td>3.2</td>
<td>1.1</td>
<td>6.9</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Correlations

Negative correlation: the perceived ‘ease’ of the writing process and the number of characters produced.

Pearson -0.461, p < .01

The writers that perceive the task being difficult need significantly more time to write a tweet.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>30</td>
<td>8‘48”</td>
<td>1’18”</td>
</tr>
<tr>
<td>difficult</td>
<td>31</td>
<td>7‘25”</td>
<td>1’18”</td>
</tr>
</tbody>
</table>
Mariëlle Leijten  
Flanders Research Foundation (FWO)  
University of Antwerp – marielle.leijten@ua.ac.be

Luuk Van Waes  
University of Antwerp  
Luuk.vanwaes@ua.ac.be

**Case | introduction**

<table>
<thead>
<tr>
<th></th>
<th>Rita</th>
<th>Ines</th>
<th>Peter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of characters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- final text</td>
<td>140</td>
<td>131</td>
<td>116</td>
</tr>
<tr>
<td>- process</td>
<td>321</td>
<td>542</td>
<td>240</td>
</tr>
<tr>
<td>total time</td>
<td>6'57</td>
<td>7'45</td>
<td>8'14</td>
</tr>
<tr>
<td>- writing time</td>
<td>4'59</td>
<td>6'09</td>
<td>5'15</td>
</tr>
</tbody>
</table>

**Case | sources**

**Case 1: Rita**

**Case 2: Ines (not familiar with twitter)**

**Distribution of activities**

**Case 1: Rita**

**Distribution of activities**

**Case 2: Ines (not familiar with twitter)**

**Distribution of activities**
**Case 3 (familiar with twitter)**

**Relations between sources**

**Relations between sources**

**Relations between sources**

**Cases Writing from multiple (digital) sources**

**Method**

- **participant**: professional writer (technical communication)
- **tasks**: daily work (proposals, reports)
- **focus**: process of writing multiple digital sources
- **duration**: 3:56:26 (three sessions)
- **analysis**: network analysis (pajek)
Case study | concise proposal

- Duration writing process 3:56.26
  - 3:04.55
  - 0:34.45
  - 0:17.26
- 131 focus events
- 10 type of documents
  - main document
  - other word documents
  - internet
  - spreadsheets
  - presentations
  - ...

Case study | elaborate proposal

- Duration writing process 9:08:02
  - 5:54:08
  - 0:39.20
  - 1:38:40
  - 0:34.45
  - 0:41.09
- 283 focus events
- 10 type of documents (based on 28 subcategories)
  - main document
  - other word documents
  - internet
  - spreadsheets
  - presentations
  - ...

Current focus

Current developments

- XML-standardization of the logfiles
- Aggregation of level of analysis
  - key/character level
  - digraph level
  - word level
  - sentence level

Collaboration: University College Ghent: Véronique Hoste & Lieve Macken
Summerschool Writing Process Research 2011: Keystroke logging & Eye tracking
Wednesday September 7 - Introduction to keystroke logging

More information

Mariëlle Leijten, Flanders Research Foundation, Belgium
marielle.leijten@ua.ac.be ~ www.ua.ac.be/marielle.leijten

Luuk Van Waes, University of Antwerp, Belgium
luuk.vanwaes@ua.ac.be ~ www.ua.ac.be/luuk.vanwaes

www.writingpro.eu
www.inputlog.net
www.jowr.org