Pro-TEXT: Annotated Corpus of Keystroke Logs

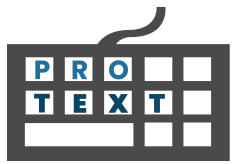
Aleksandra Miletić[1], Georgeta Cislaru[2], Christophe Benzitoun[3] and Santiago Herrera-Yanez[1]

1 Paris 3 Sorbonne Nouvelle University

2 Paris Nanterre University

3 Lorraine University

Language Resources and Evaluation Conference, June 21-23 2022



Dynamics of the Writing Process

Evolution of a text:

afin d'aborder un projet de l'Université de Poitiers
in order to address a project of the University of Poitiers

Dynamics of the Writing Process

Evolution of a text:

afin d'aborder un projet de l'Université de Poitiers
in order to address a project of the University of Poitiers

afin d'aborder un projet ~~de l'Université de Poitiers~~
in order to address a project ~~of the University of Poitiers~~

Dynamics of the Writing Process

Evolution of a text:

afin d'aborder un projet de l'Université de Poitiers
in order to address a project of the University of Poitiers

afin d'aborder un projet ~~de l'Université de Poitiers~~
in order to address a project ~~of the University of Poitiers~~

afin d'aborder un projet **songé par l'Universitté**
*in order to address a project **conceived by the Universitty***

Dynamics of the Writing Process

Evolution of a text:

afin d'aborder un projet de l'Université de Poitiers
in order to address a project of the University of Poitiers

afin d'aborder un projet ~~de l'Université de Poitiers~~
in order to address a project ~~of the University of Poitiers~~

afin d'aborder un projet **songé par l'Universitté**
*in order to address a project **conceived by the Universitty***

afin d'aborder un projet songé par l'Universit~~t~~é
in order to address a project conceived by the Universit~~t~~y

Dynamics of the Writing Process

Evolution of a text:

afin d'aborder un projet de l'Université de Poitiers
in order to address a project of the University of Poitiers

afin d'aborder un projet ~~de l'Université de Poitiers~~
in order to address a project ~~of the University of Poitiers~~

afin d'aborder un projet **songé par l'Universitté**
*in order to address a project **conceived by the Universitty***

afin d'aborder un projet songé par l'Universitté
in order to address a project conceived by the Universitty

afin d'aborder un projet songé par l'Université **de Poitiers**
*in order to address a project conceived by the University **of Poitiers***

Dynamics of the Writing Process

Evolution of a text:

afin d'aborder un projet de l'Université de Poitiers
in order to address a project of the University of Poitiers

afin d'aborder un projet ~~de l'Université de Poitiers~~
in order to address a project ~~of the University of Poitiers~~

afin d'aborder un projet **songé par l'Universitté**
*in order to address a project **conceived by the Universitty***

afin d'aborder un projet songé par l'Universitté
in order to address a project conceived by the Universitty

afin d'aborder un projet songé par l'Université **de Poitiers**
*in order to address a project conceived by the University **of Poitiers***

afin d'aborder un projet songé par ~~l'Université de Poitiers~~
in order to address a project conceived by ~~the University of Poitiers~~

Dynamics of the Writing Process

Evolution of a text:

afin d'aborder un projet de l'Université de Poitiers
in order to address a project of the University of Poitiers

afin d'aborder un projet ~~de l'Université de Poitiers~~
in order to address a project ~~of the University of Poitiers~~

afin d'aborder un projet **songé par l'Universitté**
*in order to address a project **conceived by the Universitty***

afin d'aborder un projet songé par l'Universitt~~é~~
in order to address a project conceived by the Universitt~~y~~

afin d'aborder un projet songé par l'Université **de Poitiers**
*in order to address a project conceived by the University **of Poitiers***

afin d'aborder un projet songé par ~~l'Université de Poitiers~~
in order to address a project conceived by ~~the University of Poitiers~~

afin d'aborder un projet songé par **notre Université**
*in order to address a project conceived by **our University***

Dynamics of the Writing Process

Keystroke logs: recordings of the writing process executed through a keyboard using dedicated software (Leijten and Van Waes, 2006; Strömqvist and Malmsten, 1998; Carl, 2012)

- Character additions, deletions, substitutions
- Behavioural information (pause length, production speed)
- $\rightarrow$ Study of writing itself and of underlying cognitive processes

Linguistic annotation?

- Available for English and Dutch as part of Inputlog (Leijten et al., 2015)
- Remains rare, typically POS-tagging (Serbina et al., 2015; Carl et al., 2011)
- Highly relevant, cf. (Serbina et al., 2017)

Outline

Outline

Pro-TEXT Project: Scope and Goals

Pro-TEXT: linguistic analysis of the textualization process, i.e. the real-time progressive construction of a text

- Funded by the French National Research Agency (ANR-18-CE23-0024-01)
- Interdisciplinary project:
  - **psycholinguistic experiments**: T. Olive, S. Bouriga, D. Chesnet, C. Perret, J. Pylouster and C. Bordes (CERCA, Poitiers University)
  - **linguistic description**: G. Cislaru, S. Fleury, F. Lefeuvre, D. Legallois, A. Boyer, Q. Feltgen and A. Miletic (CLESTHIA, Paris 3 University); C. Benzitoun and M. Dargnat (ATILF, Lorraine University); S. Vandaele (University of Montreal)
  - **machine learning modelling**: G. Cabanes, T. Charnois, N. Grozavu, J. Le Roux, P. Rastin, N. Rogovschi and N. Tomeh (LIPN, Paris 13 University)

Pro-TEXT Project: Scope and Goals

Pro-TEXT: linguistic analysis of the textualization process, i.e. the real-time progressive construction of a text

- Funded by the French National Research Agency (ANR-18-CE23-0024-01)
- Interdisciplinary project:
  - **psycholinguistic experiments**: T. Olive, S. Bouriga, D. Chesnet, C. Perret, J. Pylouster and C. Bordes (CERCA, Poitiers University)
  - **linguistic description**: G. Cislaru, S. Fleury, F. Lefeuvre, D. Legallois, A. Boyer, Q. Feltgen and A. Miletic (CLESTHIA, Paris 3 University); C. Benzitoun and M. Dargnat (ATILF, Lorraine University); S. Vandaele (University of Montreal)
  - **machine learning modelling**: G. Cabanes, T. Charnois, N. Grozavu, J. Le Roux, P. Rastin, N. Rogovschi and N. Tomeh (LIPN, Paris 13 University)

$\rightarrow$ Produce a corpus suitable for this type of research

Outline

Pro-TEXT Corpus

Collected Pro-TEXT Corpus:

Subcorpus	Texts	Words	Writers	Genre
Academic	26	70464	MA students	mini-thesis in linguistics
Professional	10	34504	social workers	reports on child protection
Experimental	165	63533	BA students	essays on different subjects
Children	183	20306	pupils (3^{rd}-6^{th} grade)	narrative texts and essays
Translation	38	13682	BA students	EN-FR translation of medical texts and original texts in FR
Total	422	202489	-	-

Recorded using Inputlog (Leijten and Van Waes, 2006) and Scriptlog (Strömqvist and Malmsten, 1998)

Outline

General Annotation Principles

Goal: maximize the utility of the corpus while respecting time constraints

- French Treebank tagsets (POS tags and dependency syntax) (Candito et al., 2009)
- Using automatic pre-annotation to facilitate work of human annotators
- Manual validation to ensure annotation quality
- Agile annotation (Voormann and Gut, 2008) to ensure manual annotation coherence
- Annotating **all** produced content (not only final texts)
  - $\rightarrow$ intermediate versions of each text

Intermediate Versions of a Text

Intermediate versions:

afin d'aborder un projet de l'Université de Poitiers
in order to address a project of the University of Poitiers

afin d'aborder un projet ~~de l'Université de Poitiers~~
in order to address a project ~~of the University of Poitiers~~

afin d'aborder un projet **songé par l'Universitté**
*in order to address a project **conceived by the Universitty***

afin d'aborder un projet songé par l'Universitté
in order to address a project conceived by the Universitty

afin d'aborder un projet songé par l'Université **de Poitiers**
*in order to address a project conceived by the University **of Poitiers***

afin d'aborder un projet songé par ~~l'Université de Poitiers~~
in order to address a project conceived by ~~the University of Poitiers~~

afin d'aborder un projet songé par **notre Université**
*in order to address a project conceived by **our University***

Intermediate Versions of a Text

Intermediate versions:

afin d'aborder un projet de l'Université de Poitiers → **v1**
in order to address a project of the University of Poitiers

afin d'aborder un projet ~~de l'Université de Poitiers~~ → **v2**
in order to address a project ~~of the University of Poitiers~~

afin d'aborder un projet **songé par l'Universitté** → **v3**
*in order to address a project **conceived by the Universitty***

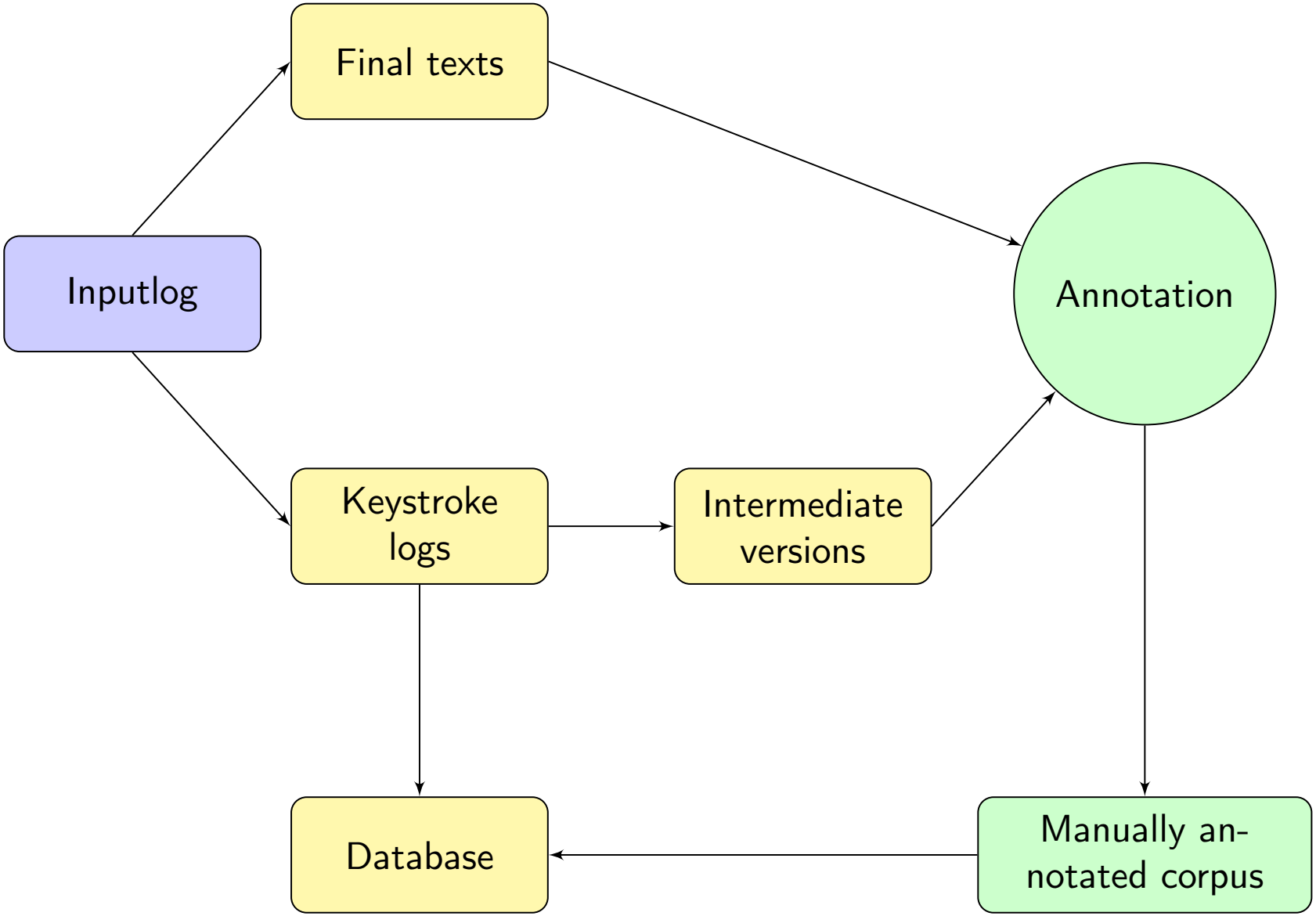
afin d'aborder un projet songé par l'Universit~~t~~é → **v4**
in order to address a project conceived by the Universit~~t~~y

afin d'aborder un projet songé par l'Université **de Poitiers** → **v5**
*in order to address a project conceived by the University **of Poitiers***

afin d'aborder un projet songé par ~~l'Université de Poitiers~~ → **v6**
in order to address a project conceived by ~~the University of Poitiers~~

afin d'aborder un projet songé par **notre Université** → **v7**
*in order to address a project conceived by **our University***

Annotation Workflow



Data Formats

Annotated intermediate versions:

- CoNLL format:

```
#text_version_id=0
#sentence_id=1
1       C'      ce      CLS     CLS     _       2       suj_impers      _       1=C__True|2='__True|
2       etait   être    V       P       _       0       root    _       3=e__True|4=t__True|5=a__True|6=i__True|7=t__True
3       en      en      P       _       _       2       p_obj   _       9=e__True|10=n__True|
4       décembre        décembre        NC      NC      _       3       prep    _       12=d__True|13=é__True|14=c__True|1
5       je      je      CLS     ADJ     _       0       x       _       21=j__True|22=e__True|
212     connaisser      _       VINF    VINF            0       _       _       24=c__True|25=o__True|26=n__True|27=n__Tru
7       ce      ce      DET     DET     _       8       det     _       35=c__True|36=e__True|
8       garçon  garçon  NC      NC      _       0       x       _       38=g__True|39=a__True|40=r__True|41=ç__True|42=o_
9       depuis  depuis  P       P       _       0       x       _       45=d__True|46=e__True|47=p__True|48=u__True|49=i_
10      la      le      DET     DET     _       11      det     _       52=l__True|53=a__True|
11      marenelle       maternelle      NC      ADJ     _       9       prep    _       55=m__True|56=a__True|57=r__True|5
12      il      il      CLS     CLS     _       0       x       _       65=i__True|66=l__True|
213     etea    _       ADV     ADV             0       _       _       68=e__True|69=t__False|70=e__False|71=a__False|

#text_version_id=1
#sentence_id=1
1       C'      ce      CLS     CLS     _       2       suj_impers      _       1=C__True|2='__True|
2       etait   être    V       P       _       0       root    _       3=e__True|4=t__True|5=a__True|6=i__True|7=t__True
3       en      en      P       _       _       2       p_obj   _       9=e__True|10=n__True|
4       décembre        décembre        NC      NC      _       3       prep    _       12=d__True|13=é__True|14=c__True|1
5       je      je      CLS     ADJ     _       0       x       _       21=j__True|22=e__True|
212     connaisser      _       VINF    VINF            0       _       _       24=c__True|25=o__True|26=n__True|27=n__Tru
7       ce      ce      DET     DET     _       8       det     _       35=c__True|36=e__True|
8       garçon  garçon  NC      NC      _       0       x       _       38=g__True|39=a__True|40=r__True|41=ç__True|42=o_
9       depuis  depuis  P       P       _       0       x       _       45=d__True|46=e__True|47=p__True|48=u__True|49=i_
10      la      le      DET     DET     _       11      det     _       52=l__True|53=a__True|
11      marenelle       maternelle      NC      ADJ     _       9       prep    _       55=m__True|56=a__True|57=r__True|5
12      il      il      CLS     CLS     _       0       x       _       65=i__True|66=l__True|
214     e       _       ADV     ADV             0       _       ***     68=e__True|
```

Data Formats

Annotated database:

- database-like CSV format:

n_event	st_time	end_time	pause	event	pos	doc_len	op	type_op	charID	tokenID	token	lemma	POS	XPOS	ms	governor	function	charStatus	tokenStatus	sentence_id
1	144098	144597		C	0	0	1	keyboard	1	-1=1|0=1|1=1|2	-1=C'|0=C'|1=C	-1=ce|0=ce|1=c	-1=CLS|0=CLS	-1=CLS|0=CLS	-1=_|0=_|1=_|2	-1=2|0=2|1=2|2	-1=suj_impers|	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
2	146235	146313	1638	'	1	1	1	keyboard	2	-1=1|0=1|1=1|2	-1=C'|0=C'|1=C	-1=ce|0=ce|1=c	-1=CLS|0=CLS	-1=CLS|0=CLS	-1=_|0=_|1=_|2	-1=2|0=2|1=2|2	-1=suj_impers|	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
3	146641	146719	328	e	2	2	1	keyboard	3	-1=2|0=2|1=2|2	-1=etait|0=etait	-1=être|0=être|	-1=V|0=V|1=V|	-1=P|0=P|1=P|	-1=_|0=_|1=_|2	-1=0|0=0|1=0|2	-1=root|0=root|	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
4	146922	146984	203	t	3	3	1	keyboard	4	-1=2|0=2|1=2|2	-1=etait|0=etait	-1=être|0=être|	-1=V|0=V|1=V|	-1=P|0=P|1=P|	-1=_|0=_|1=_|2	-1=0|0=0|1=0|2	-1=root|0=root|	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
5	147358	147421	374	a	4	4	1	keyboard	5	-1=2|0=2|1=2|2	-1=etait|0=etait	-1=être|0=être|	-1=V|0=V|1=V|	-1=P|0=P|1=P|	-1=_|0=_|1=_|2	-1=0|0=0|1=0|2	-1=root|0=root|	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
6	147577	147702	156	i	5	5	1	keyboard	6	-1=2|0=2|1=2|2	-1=etait|0=etait	-1=être|0=être|	-1=V|0=V|1=V|	-1=P|0=P|1=P|	-1=_|0=_|1=_|2	-1=0|0=0|1=0|2	-1=root|0=root|	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
7	147904	147982	202	t	6	6	1	keyboard	7	-1=2|0=2|1=2|2	-1=etait|0=etait	-1=être|0=être|	-1=V|0=V|1=V|	-1=P|0=P|1=P|	-1=_|0=_|1=_|2	-1=0|0=0|1=0|2	-1=root|0=root|	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
8	148372	148497	390	␣	7	7	1	keyboard	8	na	na	na	na	na	na	na	na	True	na	na
9	150868	150993	2371	e	8	8	1	keyboard	9	-1=3|0=3|1=3|2	-1=en|0=en|1=e	-1=en|0=en|1=e	-1=P|0=P|1=P|	-1=_|0=_|1=_|2	-1=_|0=_|1=_|2	-1=2|0=2|1=2|2	-1=p_obj|0=p_o	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
10	151102	151196	109	n	9	9	1	keyboard	10	-1=3|0=3|1=3|2	-1=en|0=en|1=e	-1=en|0=en|1=e	-1=P|0=P|1=P|	-1=_|0=_|1=_|2	-1=_|0=_|1=_|2	-1=2|0=2|1=2|2	-1=p_obj|0=p_o	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
11	151321	151414	125	␣	10	10	1	keyboard	11	na	na	na	na	na	na	na	na	True	na	na
12	151758	151836	344	d	11	11	1	keyboard	12	-1=4|0=4|1=4|2	-1=décembre|0	-1=décembre|0	-1=NC|0=NC|1	-1=NC|0=NC|1	-1=_|0=_|1=_|2	-1=3|0=3|1=3|2	-1=prep|0=prep	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
13	152959	153052	1123	é	12	12	1	keyboard	13	-1=4|0=4|1=4|2	-1=décembre|0	-1=décembre|0	-1=NC|0=NC|1	-1=NC|0=NC|1	-1=_|0=_|1=_|2	-1=3|0=3|1=3|2	-1=prep|0=prep	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
14	153364	153427	312	c	13	13	1	keyboard	14	-1=4|0=4|1=4|2	-1=décembre|0	-1=décembre|0	-1=NC|0=NC|1	-1=NC|0=NC|1	-1=_|0=_|1=_|2	-1=3|0=3|1=3|2	-1=prep|0=prep	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
15	153739	153817	312	e	14	14	1	keyboard	15	-1=4|0=4|1=4|2	-1=décembre|0	-1=décembre|0	-1=NC|0=NC|1	-1=NC|0=NC|1	-1=_|0=_|1=_|2	-1=3|0=3|1=3|2	-1=prep|0=prep	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
16	154456	154612	639	m	15	15	1	keyboard	16	-1=4|0=4|1=4|2	-1=décembre|0	-1=décembre|0	-1=NC|0=NC|1	-1=NC|0=NC|1	-1=_|0=_|1=_|2	-1=3|0=3|1=3|2	-1=prep|0=prep	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
17	157795	157920	3183	b	16	16	1	keyboard	17	-1=4|0=4|1=4|2	-1=décembre|0	-1=décembre|0	-1=NC|0=NC|1	-1=NC|0=NC|1	-1=_|0=_|1=_|2	-1=3|0=3|1=3|2	-1=prep|0=prep	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
18	158200	158263	280	r	17	17	1	keyboard	18	-1=4|0=4|1=4|2	-1=décembre|0	-1=décembre|0	-1=NC|0=NC|1	-1=NC|0=NC|1	-1=_|0=_|1=_|2	-1=3|0=3|1=3|2	-1=prep|0=prep	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
19	158419	158481	156	e	18	18	1	keyboard	19	-1=4|0=4|1=4|2	-1=décembre|0	-1=décembre|0	-1=NC|0=NC|1	-1=NC|0=NC|1	-1=_|0=_|1=_|2	-1=3|0=3|1=3|2	-1=prep|0=prep	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
20	159121	159246	640	␣	19	19	1	keyboard	20	na	na	na	na	na	na	na	na	True	na	na
21	165345	165470	6099	j	20	20	1	keyboard	21	-1=5|0=5|1=5|2	-1=je|0=je|1=je	-1=je|0=je|1=je	-1=CLS|0=CLS	-1=ADJ|0=ADJ	-1=_|0=_|1=_|2	-1=6|0=0|1=0|2	-1=suj|0=x|1=x	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
22	165642	165720	172	e	21	21	1	keyboard	22	-1=5|0=5|1=5|2	-1=je|0=je|1=je	-1=je|0=je|1=je	-1=CLS|0=CLS	-1=ADJ|0=ADJ	-1=_|0=_|1=_|2	-1=6|0=0|1=0|2	-1=suj|0=x|1=x	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
23	165938	166032	218	␣	22	22	1	keyboard	23	na	na	na	na	na	na	na	na	True	na	na
24	166250	166344	218	c	23	23	1	keyboard	24	-1=6|0=212|1=1	-1=connaissais	-1=connaître|0	-1=V|0=VINF|1	-1=V|0=VINF|1	-1=_|0=|1=|2=|	-1=0|0=0|1=0|2	-1=root|0=_|1=	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
25	166749	166843	405	o	24	24	1	keyboard	25	-1=6|0=212|1=1	-1=connaissais	-1=connaître|0	-1=V|0=VINF|1	-1=V|0=VINF|1	-1=_|0=|1=|2=|	-1=0|0=0|1=0|2	-1=root|0=_|1=	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
26	167092	167186	249	n	25	25	1	keyboard	26	-1=6|0=212|1=1	-1=connaissais	-1=connaître|0	-1=V|0=VINF|1	-1=V|0=VINF|1	-1=_|0=|1=|2=|	-1=0|0=0|1=0|2	-1=root|0=_|1=	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
27	167514	167592	328	n	26	26	1	keyboard	27	-1=6|0=212|1=1	-1=connaissais	-1=connaître|0	-1=V|0=VINF|1	-1=V|0=VINF|1	-1=_|0=|1=|2=|	-1=0|0=0|1=0|2	-1=root|0=_|1=	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
28	167794	167872	202	a	27	27	1	keyboard	28	-1=6|0=212|1=1	-1=connaissais	-1=connaître|0	-1=V|0=VINF|1	-1=V|0=VINF|1	-1=_|0=|1=|2=|	-1=0|0=0|1=0|2	-1=root|0=_|1=	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
29	168808	168949	936	i	28	28	1	keyboard	29	-1=6|0=212|1=1	-1=connaissais	-1=connaître|0	-1=V|0=VINF|1	-1=V|0=VINF|1	-1=_|0=|1=|2=|	-1=0|0=0|1=0|2	-1=root|0=_|1=	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
30	170275	170368	1326	s	29	29	1	keyboard	30	-1=6|0=212|1=1	-1=connaissais	-1=connaître|0	-1=V|0=VINF|1	-1=V|0=VINF|1	-1=_|0=|1=|2=|	-1=0|0=0|1=0|2	-1=root|0=_|1=	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
31	171070	171164	702	s	30	30	1	keyboard	31	-1=6|0=212|1=1	-1=connaissais	-1=connaître|0	-1=V|0=VINF|1	-1=V|0=VINF|1	-1=_|0=|1=|2=|	-1=0|0=0|1=0|2	-1=root|0=_|1=	True	-1=_|0=_|1=_|2	-1=1|0=1|1=1|2
32	171414	171507	250	e	31	31	1	keyboard	32	0=212|1=212|2	0=connaisserl	0=_|1=_|2=_|3	0=VINF|1=VIN	0=VINF|1=VIN	0=|1=|2=|3=|4	0=0|1=0|2=0|3	0=_|1=_|2=_|3	False	0=_|1=_|2=_|3	0=1|1=1|2=1|3=
33	171663	171757	156	r	32	32	1	keyboard	33	0=212|1=212|2	0=connaisserl	0=_|1=_|2=_|3	0=VINF|1=VIN	0=VINF|1=VIN	0=|1=|2=|3=|4	0=0|1=0|2=0|3	0=_|1=_|2=_|3	False	0=_|1=_|2=_|3	0=1|1=1|2=1|3=
34	171928	172022	171	␣	33	33	1	keyboard	34	na	na	na	na	na	na	na	na	True	na	na

Outline

Results

Annotated Final Texts:

Subcorpus	Texts	Sentences	Tokens	Tok/sent	Lemmas	Types
Children	120	440	11873	27.0	1171	224
Experimental	15	149	5719	38.4	1024	1428
Translation	10	149	3675	24.7	626	887
Academic	2	474	8879	18.7	1523	2013
Professional	0	0	0	0	0	0
TOTAL	147	1212	30146	24.9	3062	5184

Results

Annotated Intermediate Versions:

48 texts

	Sentences	Tokens	Lemmas	Types
Final texts	71	4319	632	1074
Intermediate versions	4621	128518	693	1767

Download from `https://pro-text.huma-num.fr/ressources/` under **CC BY-NC-SA 4.0** license

Outline

Future Work

Corpus annotation:

- Complete the annotation of intermediate versions
- Expand the annotation of the whole corpus: bootstrapping?

Corpus analysis:

- Linguistic profiling of writing bursts
- Relationship between linguistic structure and pauses
- Behaviour of syntactic dependencies with respect to behavioural factors
- Modelling interactions between behavioural and linguistic information using machine learning techniques

References I

Marie Candito, Benoît Crabbé, and Mathieu Falco. Dépendances syntaxiques de surface pour le français, 2009.

Michael Carl. Translog-ii: a program for recording user activity data for empirical reading and writing research. In LREC, volume 12, pages 4108–4112, 2012.

Michael Carl, Barbara Dragsted, Jakob Elming, Daniel Hardt, and Arnt Lykke Jakobsen. The process of post-editing: A pilot study. Copenhagen Studies in Language, 41:131–142, 2011.

Mariëlle Leijten and Luuk Van Waes. Inputlog: New perspectives on the logging of on-line writing processes in a windows environment. In Computer key-stroke logging and writing, pages 73–93. Brill, 2006.

Mariëlle Leijten, Luuk Van Waes, and Eric Van Horenbeeck. Writing(s) at the Crossroads: The Process-Product Interface, chapter Analyzing writing process data: A linguistic perspective, pages 277–302. John Benjamins Publishing Company, 2015.

Tatiana Serbina, Paula Niemietz, Matthias Fricke, Philipp Meisen, and Stella Neumann. Part of speech annotation of intermediate versions in the keystroke logged translation corpus. In Proceedings of the 9th LinguiLtic Annotation Workshop, pages 102–111, 2015.

References II

Tatiana Serbina, Sven Hintzen, Paula Niemietz, and Stella Neumann. Empirical modelling of translation and interpreting, volume 3, chapter Changes of word class during translation–Insights from a combined analysis of corpus, keystroke logging and eye-tracking data, pages 177–208. Language Science Press, Berlin, 2017.

Sven Strömqvist and Lars Malmsten. Sriptlog pro 1.04: User's manual. Technical report, University of Göteborg, 1998.

Holger Voormann and Ulrike Gut. Agile corpus creation. Corpus Linguistics and Linguistic Theory, 4(2):235–251, 2008.