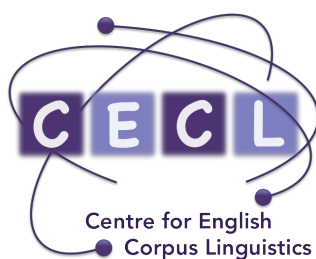


CECL Papers 2

Annotating the use of online writing resources in a video corpus of written process data in ELAN

Annotation manual version 1.1

Samantha Laporte and Gaëtanelle Gilquin



2018

Table of Contents

| | |
|---|-----------|
| 1. Introduction: research context and objectives | 3 |
| 2. Annotation protocol | 4 |
| 2.1. Overview of the annotation scheme..... | 4 |
| 2.2. Technical definition of the tiers in ELAN | 4 |
| 2.3. Tiers: definition and description | 6 |
| 2.3.1. CONSULTATION_UNIT | 6 |
| 2.3.2. TOOL | 7 |
| 2.3.3. INFORMATION_TYPE..... | 8 |
| 2.3.4. QUERY | 9 |
| 2.3.5. EMPTY_QUERY | 10 |
| 2.3.6. INDIRECT_QUERY | 10 |
| 2.3.7. INDIRECT_ACCESS | 11 |
| 2.3.8. EFFECT_UNIT | 11 |
| 2.3.9. RESULT..... | 12 |
| 2.3.10. COMMENT | 13 |
| 2.3.11. RECORDING_PROBLEM..... | 13 |
| 2.3.12. PARALLEL_USE | 14 |
| 3. Conclusion | 14 |
| References..... | 14 |

Citing this document: Laporte, S. and Gilquin, G. 2018. Annotating the use of online writing resources in a video corpus of written process data in ELAN. Annotation manual version 1.1. CECL Papers, vol. 2. Louvain-la-Neuve: Université catholique de Louvain.

1. Introduction: research context and objectives

Until recently, writing has been almost exclusively studied through its final product, i.e. the written text. The process of writing, by contrast, that is, the different steps leading to the final product, has not been much investigated. This is mainly due to the fact that this process is difficult to observe. Typically, traces of editing are no longer visible when a text reaches its audience. New techniques like keylogging (the recording of the keys struck on a computer keyboard) and screencasting (the recording of a computer screen on video), however, have made it possible to capture the writing process like never before, including editing, but also the use of online resources like dictionaries, databases or secondary sources. So far, these techniques have mostly been applied to professional writing, including translation and newswriting (Ehrensberger-Dow & Perrin, 2013; Perrin, 2013). Applications to (first or second) language acquisition are still rare (see Séror, 2013 for an exception).

The PROCEED (Process Corpus of English in Education) project aims to capture the writing process of foreign learners of English via keylogging and screencasting. Its compilation started in February 2017 at the University of Louvain, Belgium. Learners' computer activity was recorded while they were writing a short argumentative essay on one of several set topics. Inputlog (Leijten & Van Waes, 2013) was used to record keyboard activity, while OBS (Open Broadcaster Software) recorded screen activity. PROCEED includes three types of data: the written texts (Word documents), the keystroke logging data (log files) and the screencasts (videos). To make the videos amenable to corpus analysis, it is necessary to describe their contents in a format that can be exploited by text retrieval software. This can be done through annotation, which provides written information about what can be seen on the videos.

The annotation scheme detailed here has been developed within the framework of PROCEED to analyze students' use of online writing resources (dictionaries, corpus interfaces, etc.) during the writing process. It allows for the investigation of the time spent using tools, the number of searches carried out, the tools that students use, whether tools are accessed directly or indirectly, for what query and what purpose tools are used, whether tools are used in combination, and whether the use of tools actually helps students improve their texts.

The software tool used to annotate the videos is ELAN¹ (Wittenburg et al., 2006), an open source video and audio annotation program developed at the Max Planck Institute for Psycholinguistics in Nijmegen. As a highly complex and flexible tool, it enables the creation of annotation layers, or 'tiers', to capture the different features that are of interest for the annotator. Annotation tiers can be related to each other in different ways so as to form a hierarchy of interconnected tiers. The annotation files are created and stored in the ELAN Annotation Format (EAF), and the annotations can be queried and analyzed within the ELAN interface, or exported in multiple formats to be exploited with other software tools for analysis.

While developed in a specific context with the primary aim of annotating the video component of PROCEED to analyze the use of online resources by students writing in L2 English, the present scheme aims to be applicable to any kind of annotation of the use of writing tools in all types of writing situations by any type of user, including professional writers or translators.

¹ Available at <https://tla.mpi.nl/tools/tla-tools/elan/>.

2. Annotation protocol

2.1. Overview of the annotation scheme

To capture the many dimensions of the use of writing tools mentioned above, an ELAN annotation template of 12 tiers has been created. Table 1 provides a succinct overview of the set of tiers and a description of what each tier designates in the annotation. The annotation of some of these tiers is constrained by a predefined list of annotation values (called 'Controlled Vocabulary' in ELAN), while others can be freely annotated. An asterisk indicates those tiers that have a predefined list of possible values.

| Tier name | Description |
|--------------------|--|
| CONSULTATION_UNIT* | Each search, defined as each time one or several tools are used to answer a specific question |
| TOOL* | Tool that is used (e.g. Macmillan dictionary, Thesaurus.com) |
| INFORMATION_TYPE* | Type of information that is looked up (e.g. spelling, synonym) |
| QUERY | Exact query or queries in a tool (e.g. 'comitted') |
| EMPTY_QUERY | Indicates that the query did not return any result in the tool (e.g. Ozdic does not have an entry for <i>much</i>) |
| INDIRECT_QUERY | Indicates when a query is indirect, i.e. when the search results of a tool prompt a new query within the same tool (e.g. one of the synonyms provided by a thesaurus is clicked on to find further synonyms) |
| INDIRECT_ACCESS | Indicates when a tool is accessed indirectly (e.g. tool accessed after a query in Google) |
| EFFECT_UNIT* | Effect of the entire search on the student's written production (e.g. 'Positive', 'Negative', 'None') |
| RESULT* | What was written or revised after the search (e.g. 'comitted > committed') |
| COMMENT | Any additional comment/observation deemed relevant |
| RECORDING_PROBLEM | Indicates technical problem with the video that prevents video annotation |
| PARALLEL_USE | Indicates the use of the Internet and Word windows next to each other on the screen |

Table 1. Overview of the annotation scheme

2.2. Technical definition of the tiers in ELAN

In ELAN, each tier of an ELAN annotation scheme needs to be configured for a number of settings. The creation of a tier in ELAN is a two-step process. First, a **tier type** has to be defined, and only then can an annotation **tier** proper be created, as a tier has to be assigned a tier type. While a single tier type can in theory be used to then create multiple annotation tiers, in the present annotation scheme, each tier has its own tier type that carries the same name.

A **tier type** is defined according to a number of dimensions (cf. Figure 1). The dimensions specified within the framework of the present work are:

- **Type Name:** given name for a tier type.
- **Stereotype:** constraints that apply to a tier. In effect, this constrains the relationship between a parent tier and a child tier by, for example, stipulating whether annotation spans of a child tier have to fully map onto those of its parent tier or whether there can be subdivisions in the child tier. The two types of stereotypes used here are:
 - o **Included In:** annotations in the child tier have to fall within annotation spans of the parent tier, but within this span, any number of time-aligned annotation units for the targeted phenomenon can occur.
 - o **Symbolic Association:** there is a one-to-one mapping between the annotation spans in the parent and child tier. The time-alignment is therefore the same as that of the parent tier.
- **Use Controlled Vocabulary:** if applicable, predefined list of values for the annotation of a tier. This Controlled Vocabulary has to be created before adding the tier type (cf. ELAN manual, available at <https://tla.mpi.nl/tools/tla-tools/elan/>).

Once a tier type is created, a **tier** is added, which, like the tier type, has to be defined according to a number of dimensions (cf. Figure 2). The dimensions specified in this work are:

- **Tier Name:** given name for a tier that here corresponds exactly to the name of its corresponding tier type.
- **Parent Tier:** if applicable, specifies the parent tier of a tier.
- **Tier Type:** defines the tier type, which here is the tier type that has been defined for this particular tier.

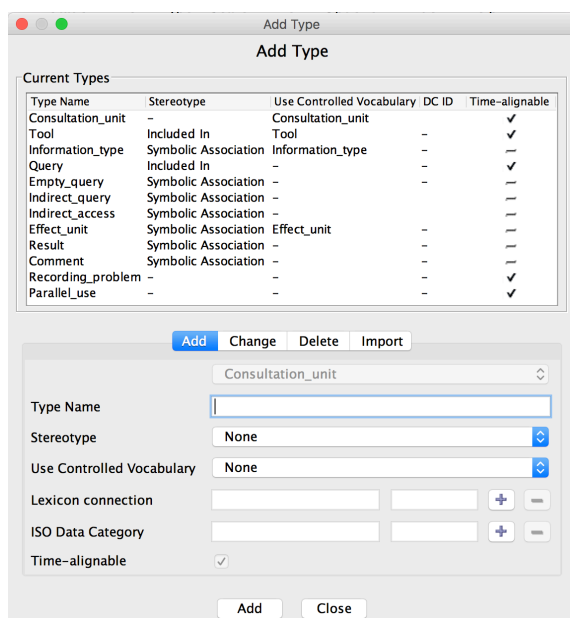


Figure 1. Defining a tier type

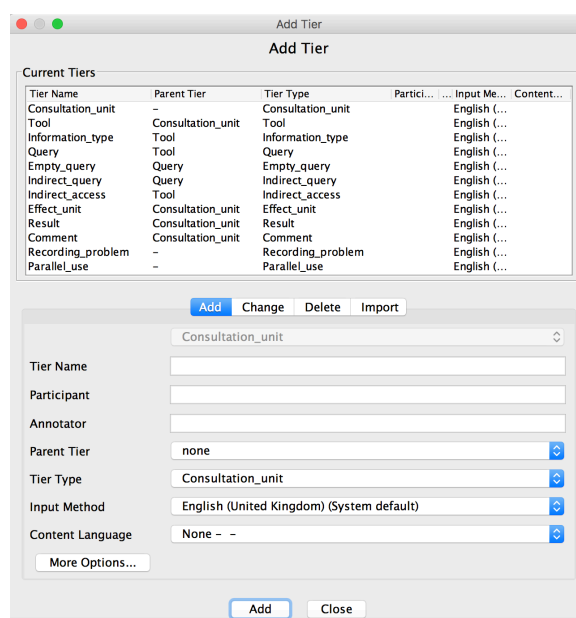


Figure 2. Adding a tier

Table 2 shows how each tier and its corresponding tier type have been defined according to these dimensions, and visually represents the parent-child relations between tiers.

| Tier | | Tier type | | |
|-------------------|-------------------|-------------------|----------------------|-----------------------|
| Tier name | Parent tier | Type name | Stereotype | Controlled Vocabulary |
| CONSULTATION_UNIT | - | Consultation_unit | - | Yes |
| TOOL | CONSULTATION_UNIT | Tool | Included in | Yes |
| INFORMATION_TYPE | TOOL | Information_type | Symbolic association | Yes |
| QUERY | TOOL | Query | Included in | No |
| EMPTY_QUERY | QUERY | Empty_query | Symbolic association | No |
| INDIRECT_QUERY | QUERY | Access_query | Symbolic association | No |
| INDIRECT_ACCESS | TOOL | Access_tool | Symbolic association | No |
| EFFECT_UNIT | CONSULTATION_UNIT | Effect_unit | Symbolic association | Yes |
| RESULT | CONSULTATION_UNIT | Result | Symbolic association | No |
| COMMENT | CONSULTATION_UNIT | Comment | Symbolic association | No |
| RECORDING_PROBLEM | - | Recording_problem | - | No |
| PARALLEL_USE | - | Parallel_use | - | No |

Table 2. Technical definition of each tier and tier type

2.3. Tiers: definition and description

2.3.1. CONSULTATION_UNIT

The CONSULTATION_UNIT tier serves to indicate the timespans on a video that cover one **search**, which is defined as each time one or several tools are used to answer a specific question. A CONSULTATION UNIT may therefore cover the interrupted use of a single tool, several queries within a tool, or the use of several tools, e.g. when a user looks up some information, goes back to their text, and then returns to the same tool or uses another tool to check some additional information related to the same search.

The different types of CONSULTATION UNITS are annotated via a predefined list of values in a Controlled Vocabulary, which includes the following values:

- **Single:** single timespan with the use of a single tool for one specific question. Note that this may include the use of a search engine (typically Google) to look for the tool (e.g. the user keys in 'Reverso' in Google to find and access the tool) before the actual use of the tool for a language-related query. While technically, two resources are used, only the writing tool serves a writing-related purpose, which is why the CONSULTATION UNIT is annotated as the use of a single tool. On the INFORMATION_TYPE tier, the use of a search engine to find another tool is annotated as 'Other/content'.
- **Cluster_single_tool:** interrupted use of a single tool (e.g. the user consults an online dictionary, goes back to their text, and then returns to the same dictionary to check the information displayed or to perform a new query related to the same question).
- **Cluster_combination:** interrupted or uninterrupted use of two or more different tools in combination with each other for one search (e.g. the user looks up the translation of a French word or expression in a bilingual dictionary and then uses a monolingual dictionary to check the use of the translation found).

- **Other_use**: use of an online resource for non-language-related searches, generally to look up some content-related information (typically, the use of Google, followed by the consultation of other webpages). Note that, as specified above, when a search engine is used to look for another tool (e.g. a collocation dictionary), it is included in the CONSULTATION UNIT that covers the subsequent use of the tool and the CONSULTATION UNIT is annotated as 'Single'.
- **?**: used in case of doubt in annotation.

In addition, timespans that should not be considered in the analysis of the video are also annotated on this tier via the 'TRIM' label:

- **TRIM**: this value has been added to select timespans that are not devoted to the writing process at the beginning and end of each video. This subsequently enables an accurate calculation of the time spent writing. Within the frame of the PROCEED data, the beginning of each video is annotated as 'TRIM' until the user starts typing the quote they have selected as topic. At the end of the video, the 'TRIM' span starts when the user no longer performs any writing-related activity, including reading over their text.

2.3.2. TOOL

The TOOL tier is a child tier of CONSULTATION_UNIT and indicates the tool or tools that are used in a CONSULTATION UNIT.

The annotation of the TOOL tier allows the span covered by CONSULTATION_UNIT to be divided into time-aligned subunits covering the use of different tools, or the interrupted use of a single tool.

The annotation is constrained by a Controlled Vocabulary that lists the different tools that are encountered, including:

- | | |
|-------------------------------|-----------------------------------|
| ○ Google_dictionary | ○ Ozdic |
| ○ Google_search | ○ Reverso_Context_ENG_FR |
| ○ Larousse_ENG_FR | ○ Reverso_Context_FR_ENG |
| ○ Larousse_FR_ENG | ○ Reverso_ENG_FR |
| ○ Linguee | ○ Reverso_FR_ENG |
| ○ Macmillan_dictionary | ○ Thesaurus |
| ○ Macmillan_thesaurus | ○ Wordreference_dictionary |
| ○ OALD_dictionary | ○ ... |

This list is not exhaustive, as the Controlled Vocabulary is edited and supplemented with the names of new tools encountered in the data as annotation goes along. It is therefore important to save the new ELAN annotation template with the new Controlled Vocabulary each time a new tool is added to ensure that the same tool is consistently annotated in the same way across data files. Note that for translation tools, the direction of translation is specified in the Controlled Vocabulary (e.g. Reverso_FR_ENG vs. Reverso_ENG_FR) and that for resources that

are made of different components that represent different types of tools, the component in question is specified (e.g. Macmillan_dictionary vs. Macmillan_thesaurus).

Multiple queries

When multiple queries are carried out during the *uninterrupted* use of a tool and the INFORMATION TYPE for each query is the same, these are covered by one unit at the TOOL level. However, when the INFORMATION TYPE for the queries is different, two (or more) units are distinguished on the TOOL tier so as to be able to account for the difference in INFORMATION TYPE at that level of annotation (e.g. a user looks up the word 'urge' in Reverso_FR_ENG to check how it is used but because the query is by mistake carried out in the French-English direction, the query does not return any result. The user then looks up the translation of 'pulsion').

Technical note if the data is to be exploited in Excel afterwards: when the TOOL tier is subdivided into different units, it is advisable for the left boundary of the first subunit to come slightly after that of its mother CONSULTATION_UNIT (one millisecond suffices). When data is extracted in Excel, it is by default sorted on the timestamp that indicates the begin time of any annotation unit. Slightly offsetting the left boundary of the first subunit on the TOOL tier allows all rows of these subunits to come below the row of their mother CONSULTATION_UNIT, which is far more convenient for subsequent data analysis.

2.3.3. INFORMATION_TYPE

The INFORMATION_TYPE tier indicates the kind of information that is looked up for each use of a tool. INFORMATION_TYPE is a child tier of TOOL and has a one-to-one relation with TOOL, which means that each annotation unit of INFORMATION_TYPE corresponds exactly to an annotation unit of TOOL.

The annotation is constrained by a Controlled Vocabulary that lists the different kinds of information distinguished here, which are the following:

- **Use/definition:** when a user verifies the meaning or use of a search term.
- **Translation:** when a user looks for a translation of a search term.
- **Synonym:** when a user looks for a synonym of a search term.
- **Collocation:** when a user looks for a collocation of a search term.
- **Spelling:** when a user verifies the spelling of a search term.
- **Grammar:** when a user looks for grammatical information.
- **Other/content:** when the user uses a tool for a non-writing-related purpose (e.g. to understand a quote in the essay guidelines, to look up general information, to look for a tool).
- **?:** used in case of doubt in annotation.

By default, INFORMATION TYPE is annotated for the primary function of the tool (e.g. 'Synonym' when a thesaurus is used, 'Translation' when a bilingual dictionary is used), unless the tool is clearly used for a different purpose.

2.3.4. QUERY

The QUERY tier indicates the exact query or queries that a user performs in a tool. QUERY is a child tier of TOOL and allows the span covered by TOOL to be divided into time-aligned subunits covering the queries carried out in the case of multiple queries.

The tier is freely annotated (i.e. there is no Controlled Vocabulary). The following guidelines specify how to annotate this tier in a number of specific cases:

Interrupted use of a tool

When, in the case of an interrupted use of a tool, no new query is carried out, the QUERY field is left blank in the second (or further) instance of the use of the tool, as the query appears in the first use of the tool (e.g. a user looks up 'richesse' in a bilingual dictionary, goes back to their text, and then returns to the tool to check the spelling of 'wealth'). By contrast, in the case of an interrupted use of a tool where the user performs a new query, the new query must be indicated.

Indirect tool access

The QUERY field is left blank when the user accesses a tool indirectly via a search engine as the query appears in the Google search right before this (e.g. if the Google query 'equality synonym' leads the user to the Thesaurus.com results page for 'equality', the query is left blank for the use of Thesaurus.com).

Indirect query

When a user carries out a new query in a tool by clicking on a term in a results page (e.g. a user looks up synonyms for 'argue' and clicks on the term 'contend' in the results), this is considered a query in its own right and is thus time-aligned as a single unit on the QUERY tier. However, since this is an indirect query prompted by the tool rather than decided by the user, this is indicated on the INDIRECT_QUERY tier (cf. Section 2.3.6).

Incremental search

When a query is partly arrived at through the incremental search feature of a tool, the part of the query that the user has keyed in is followed by what has been suggested via the incremental search in parentheses (e.g. 'montre (de luxe)' when a user keys in 'montre' in Linguee, which suggests 'montre de luxe'). In addition, 'Incremental search' is also annotated on the INDIRECT_QUERY tier (cf. Section 2.3.6).

Technical note if the data is to be exploited in Excel afterwards: when the QUERY tier is subdivided into different units, it is advisable for the left boundary of the first subunit to come slightly after that of its mother TOOL (one millisecond suffices). When data is extracted in Excel, it is by default sorted on the timestamp that indicates the begin time of any annotation unit. Slightly offsetting the left boundary of the first subunit on the QUERY tier allows all rows of these subunits to come below the row of their mother TOOL, which is far more convenient for subsequent data analysis.

2.3.5. EMPTY_QUERY

The EMPTY_QUERY tier serves to indicate that a query does not return any result in the tool (e.g. a user looks up 'much' in Ozdic, but Ozdic does not have any entry for 'much'). This includes cases where the tool returns information about another word (e.g. a user keys in 'urge' in a dictionary, but the tool returns results for 'urgent'), or when the tool suggests close matches (e.g. the tool suggests 'commit' and 'comity' when a user queries 'comit').

EMPTY_QUERY is a child tier of QUERY and has a one-to-one relation with QUERY, which means that each annotation unit of EMPTY_QUERY corresponds exactly to an annotation unit of QUERY.

The EMPTY_QUERY tier is freely annotated (i.e. there is no Controlled Vocabulary).

When a query does not return any result, the tier is annotated with the term 'Empty', which can be complemented with additional information, like the close match(es) suggested by the tool. By default, this tier is left blank when a query returns a result.

Note that when a search only consists of one or several queries that do not return any result, the 'Empty' annotation on the EMPTY_QUERY tier corresponds to an 'Empty' annotation on the EFFECT_UNIT tier (cf. Section 2.3.8).

2.3.6. INDIRECT_QUERY

The INDIRECT_QUERY tier serves to indicate when a query is indirect, i.e. when a query in a tool is prompted by results provided by that tool rather than initiated by the user (e.g. one of the synonyms provided by a thesaurus is clicked on to find further synonyms).

The INDIRECT_QUERY tier is a child tier of QUERY and has a one-to-one relation with QUERY, which means that each annotation unit of INDIRECT_QUERY corresponds exactly to an annotation unit of QUERY.

The INDIRECT_QUERY tier is freely annotated (i.e. there is no Controlled Vocabulary).

When a query is indirect, the tier is annotated with the term 'Indirect query', followed by any additional information deemed relevant between parentheses (e.g. 'Indirect query (clicks on 'secretary')'). By default, this tier is left blank when a query is direct.

Incremental search

When a query is partly carried out via an incremental search (e.g. a student keys in 'montre' and Linguee suggests 'montre de luxe', which the user then clicks on), this is indicated here with the label 'Incremental search'.

2.3.7. INDIRECT_ACCESS

The INDIRECT_ACCESS tier serves to indicate when access to a tool is indirect, i.e. the user accesses a tool via a general query in a search engine like Google (e.g. 'synonym of interesting'). This includes cases where a user clicks on a word in a tool and that this leads to another tool or tool component that is considered a different tool on the TOOL tier (e.g. accessing Macmillan_thesaurus by clicking on the link from an entry in Macmillan_dictionary).

The INDIRECT_ACCESS tier is a child tier of TOOL and has a one-to-one relation with TOOL, which means that each annotation unit of INDIRECT_ACCESS corresponds exactly to an annotation unit of TOOL.

The INDIRECT_ACCESS tier is freely annotated (i.e. there is no Controlled Vocabulary).

When access is indirect, the tier is annotated with the term 'Indirect', which is followed, between parentheses, by the tool from which the search originates (e.g. 'Indirect (Google)' when a tool was accessed from a Google search). If the tool is accessed indirectly for a previous query and reused for a new query, this is indicated by adding 'reuse' between parentheses instead (e.g. 'Indirect (reuse)'). Note that the 'reuse' label ignores changes in language direction within the same resource (e.g. Reverso_FR_ENG and Reverso_ENG_FR) or the use of a different component within the same tool (e.g. Macmillan_dictionary and Macmillan_thesaurus). By default, the tier is left blank when access to a tool is direct.

2.3.8. EFFECT_UNIT

The EFFECT_UNIT tier describes the effect of the search (which may cover the use of several tools) as positive, negative, or lacking, according to whether or not the search has helped the user improve their text.

The EFFECT_UNIT tier is a child tier of CONSULTATION_UNIT and has a one-to-one relation with CONSULTATION_UNIT, which means that each annotation unit of EFFECT_UNIT corresponds exactly to an annotation unit of CONSULTATION_UNIT.

The coding scheme adopted here has been adapted from that of Kotamjani et al. (2017), who investigate the use of a number of online corpus tools by a single participant and who classify the effect of the use of a tool as positive, negative or lacking. Their coding scheme has however been elaborated on to arrive at a more fine-grained analysis of the effect.

The effect of a search is annotated according to the following coding scheme:

- **Positive:** the search leads to a correct formulation or validates a correct formulation in the text (e.g. a user looks up 'soulagement' in a bilingual dictionary, and then replaces 'a feeling of relief' with 'a sense of relief').
- **Negative:** the search leads to an incorrect formulation or validates an incorrect formulation in the text (e.g. a user looks up 'poursuivre' in a bilingual dictionary and then writes 'pursue' when, given the legal context, 'sue' or 'prosecute' would have been the correct formulation). The negative effect is broken down according to whether the incorrect use is one of form, meaning, style or a combination of these (see below).

- **Both:** the search leads to a partly correct, partly incorrect formulation, or validates a partly correct, partly incorrect formulation in the text (e.g. a user looks up 'besoins de base' in a bilingual dictionary, and writes 'basics needs' instead of 'basic needs' as suggested by the dictionary). Similarly to the negative effect, this category is broken down according to whether the incorrect use is one of form, meaning, style or a combination of these (see below).
- **None:** the tool provides some information, but it is not relevant or not used properly by the user. Therefore, consultation does not help to find information that has an effect on the written production (e.g. a user looks up 'appréciable' in a bilingual dictionary in the hope of finding a translation that fits the context, but the translations provided are not helpful).
- **Empty:** the tool does not return any information (e.g. a user looks up 'paradis fiscaux' in a bilingual dictionary, but the dictionary does not return any result for this query). This label is also applied when the tool returns information about another word (e.g. a user keys in 'urge' in a dictionary, but the tool returns results for 'urgent'). Note that the effect of a consultation unit is only annotated as 'Empty' when *none* of the queries returns any result. This means that this label overlaps with the annotation on the EMPTY_QUERY tier (cf. Section 2.3.5). By contrast, however, a query annotated as 'Empty' on the EMPTY_QUERY tier does not necessarily entail an 'Empty' annotation on the EFFECT_UNIT tier, as it may be the case that another query within the same CONSULTATION_UNIT does lead to an effect.
- **Indeterminate:** used when the effect is unclear.

The annotation is constrained by a Controlled Vocabulary that follows the above coding scheme, with the categories 'Negative' and 'Both' broken down according to the type of incorrect use:

- | | |
|-------------------------------------|---------------------------------|
| ○ Positive | ○ Both_meaning |
| ○ Negative_form | ○ Both_style |
| ○ Negative_meaning | ○ Both_form&style |
| ○ Negative_style | ○ Both_form&meaning |
| ○ Negative_form&style | ○ Both_meaning&style |
| ○ Negative_form&meaning | ○ None |
| ○ Negative_meaning&style | ○ Empty |
| ○ Both_form | ○ Indeterminate |

2.3.9. RESULT

The RESULT tier reproduces what is written or revised in the text as a result of a search.

The RESULT tier is a child tier of CONSULTATION_UNIT and has a one-to-one relation with CONSULTATION_UNIT, which means that each annotation unit of RESULT corresponds exactly to an annotation unit of CONSULTATION_UNIT.

The RESULT tier is freely annotated (i.e. there is no Controlled Vocabulary).

Four different types of results are distinguished:

- **Revision:** when the user revises some of their writing as a result of the search. A revision is annotated by indicating the original and the revised text separated by the 'greater than' sign > (e.g. 'aces > access').
- **New formulation:** when the user writes some new text as a result of the search. This is indicated by the 'greater than' sign >, followed by the formulation that appears in the text (e.g. '> commit a crime').
- **No change:** when nothing is changed in the text because the search validates some (correct or incorrect) formulation. This is indicated by the equals sign =, followed by the word(s) that remain(s) that remains the same in the text (e.g. '= incident').²
- **Unrelated result:** when nothing related to the search, either before or after the search, appears in the text. This is indicated with the sign ø.

Sometimes, additional context or comments are provided in parentheses (e.g. '(pave the road) for > to' where the context of use of the preposition that is edited is specified, '> the detainee (for 'inmate' or 'prisoner')' where the comment explains what the writer means).

2.3.10. COMMENT

The COMMENT tier is a tier that lets the annotator provide any additional comment or observation about a search (e.g. additional information about errors, additional explanation about some annotation).

The COMMENT tier is a child tier of CONSULTATION_UNIT and has a one-to-one relation with CONSULTATION_UNIT, which means that each annotation unit of COMMENT corresponds exactly to an annotation unit of CONSULTATION_UNIT.

The COMMENT tier is freely annotated (i.e. there is no Controlled Vocabulary).

2.3.11. RECORDING_PROBLEM

The RECORDING_PROBLEM tier allows the annotator to identify the occurrence of a technical problem with the screencasting (e.g. freezing or lagging of the video) due to which no analysis or only a partial analysis and annotation of the video can be performed. The annotation makes it possible to exclude those sections from data analysis.

The RECORDING_PROBLEM tier is freely annotated (i.e. there is no Controlled Vocabulary).

In the event of a recording problem, the tier is annotated with the term 'Recording problem', followed by any other information deemed relevant between parentheses (e.g. 'Recording problem (slight lagging of the video that makes time-alignment slightly inaccurate)').

² If the data is to be extracted and exploited in Excel after annotation, a single inverted comma should be added before the equals sign (e.g. '= incident') to prevent Excel from treating the cells that start with an equals sign as a formula.

2.3.12. PARALLEL_USE

The PARALLEL_USE tier indicates timespans when the user displays the Internet page and their text file in parallel on their screen. This serves to account for the fact that this writing strategy biases the amount of time spent on using tools, since the user can, at the same time, write in their text file and look at the information provided by the tool.

The PARALLEL_USE tier is freely annotated (i.e. there is no Controlled Vocabulary).

When the text file and Internet page are used in parallel, the tier is annotated with the term 'Parallel use', followed by any other information deemed relevant between parentheses.

3. Conclusion

Drawn up in a spirit of replicability and comparability, the present manual has sought to provide a detailed protocol to annotate the use of online writing resources in a video corpus with the ELAN software. While developed within the framework of a project that investigates the L2 writing process, the protocol aims to be applicable in the wider context of video annotation of any use of tools, possibly with annotation software other than ELAN. It is therefore hoped that this protocol will serve as a starting point or point of comparison for researchers interested in annotating video content for the use of online resources in any writing context.

References

- Ehrensberger-Dow, M. & Perrin, D. 2013. Applying a newswriting research approach to translation. *Target* 25 (1), 77-92.
- ELAN (Version 5.2) [Computer software]. (2018, April 4). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from <https://tla.mpi.nl/tools/tla-tools/elan/>
- Kotamjani, S. S., Ommehoney, F. R. & Hussin, H. 2017. Online corpus tools in scholarly writing: A case of EFL postgraduate student. *English Language Teaching* 10 (9), 61-68.
- Leijten, M. & Van Waes, L. 2013. Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication* 30 (3), 358-392.
- Perrin, D. 2013. *The Linguistics of Newswriting*. Amsterdam: John Benjamins.
- Séror, J. 2013. Screen capture technology: A digital window into students' writing processes. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie* 39 (3), 1-16.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. 2006. ELAN: A Professional Framework for Multimodality Research. In: *Proceedings of LREC 2006*, Fifth International Conference on Language Resources and Evaluation.