NARODOWE CENTRUM NAUKI

Annotation and transcription standards for research into the process of EFL lexico-grammatical consultation in the Web 2.0 environment

Marek Molenda, University of Łódź
Anastasiia Anisimova, University of Łódź

Table of contents

Introduction

The purpose of this document is to provide clear guidelines for transcribing learner data. The usage scenario envisages studies whose aim is to describe and analyze various aspects of the use of online dictionaries and/or other language-oriented consultation sources. The framework proposed is based on the premise that data would be collected by means of screen casting (or similar to screen casting), with the caveat that the information must be at least partially recorded and interpreted by humans. Consequently, solutions such as log files, which are a fully automated way of collecting similar data, do not fall within the scope of this project.

Our transcription framework is meant to be a universal tool, thus it might be modified so as to be better adjusted to the needs of the researcher. However, in order to maintain uniformity, transparency and transferability of research results, we propose that certain elements of it remain unchanged. These elements are marked with the letter "C" ("Core"), while the optional ones are marked with an "O".

Basic premises

Our framework is based upon the interpretation of Activity Theory proposed by Geisler and Slattery (2007). These authors postulate that recording human activity on screen should be based on the following premises:

1. Human behavior is goal-oriented – thus a decision to perform a given operation is, under normal circumstances, dictated by a desire to obtain some kind of meaningful effect which would move one closer to the envisaged outcome.
2. Human behavior is hierarchical. The most basic unit are **operations** – unconsciously-performed single events, such as clicking on a button. By their very nature, they are meaningless to the researcher without proper context. The next level is **actions**. These are composed of operations, but they are conscious, and they provide context for a given operation. For instance, while typing the word "fought" in the textbox or clicking on a "Search" button are operations, together they form the action of looking up a given word in an electronic dictionary. Various actions grouped together to achieve a more general goal are referred to as **activities**. An example of an activity would be performing various actions (word-based lookup, in-text search, definition-based lookup, etc.) across multiple sources in order to find the optimal word to be used in a given context.
3. Human behavior is both external and internal. Certain processes that occur in one's minds lead to manifestable interactions with the external world. A dictionary lookup activity is, therefore, a combination of manifested and unmanifested processes.
4. Human behavior is always mediated with tools – both mental and external. Mental tools such as dictionary skills are manifested by interactions with specific external artifacts, namely consultation sources.
5. Human behavior develops over time. This aspect, related to the constant dismantling and re-building of the hierarchy of activities, actions and operations means that different subjects, when recorded on a single occasion, might perform the same tasks on various levels. For example, looking for information in a collocations dictionary might be an automatized operation for one student, while another – who has just discovered this tool – will need to perform at the level of conscious Actions to arrive at the same result.

In sum, Activity Theory provides a framework which allows one to record and interpret the phenomena related to learners' use of consultation sources based on their context, preceding and following

decisions, perceived goals and actual results. In addition, it ensures that both high-granularity and low-granularity phenomena are recorded and taken into consideration.

Basic categories

Table 1 presents basic categories used in the process of data transcription:

Name	Definition and/or example	Remarks	C/O
Operation	Single event, e.g., *click on the "search" button*; *type the word "guard" in the textbox*	Every operation should have its starting and ending time recorded. This not only makes it possible to compare the length of similar operations across subjects, but it also allows one to determine the length of actions and activities.	C
Action	Sets of operations carried out in order to achieve a single, clearly defined goal. For instance, the action of *looking for the meaning of the word "caterpillar"* might consist of the following operations: *open LDOCE* *type "caterpillar"* *click on the search button* *analyze the entry* *scroll down the entry* *scroll up* *close the window*	Action might span across various sources provided that the **aim** remains unchanged.	C
Activity	Sets of actions carried out in order to solve a given problem (lexico-grammatical information gap). For instance, *finding the most appropriate collocation of the word "cone"* which would fit in the phrase "to cone … the road" might comprise the following actions: *look for the meaning of the word "cone"* *verify the hypothesis that "cone out" would be the correct answer* *look for collocations of the word "cone"* *verify the hypothesis that "cone off" would be the right answer* *type the answer in the text*	As shown in the example, **actions** might be assigned an additional meaning (e.g., looking for new information, verification, etc.) solely in the context of **activities** which outline the global goal.	C

Table 1. Basic categories

Under optimal conditions, operations would always form consecutive actions, and actions combine to form activities in a linear way without any overlaps (Figure 1.1). However, our data shows that the subjects might interrupt a given action or activity in order to return to it later (1.2).

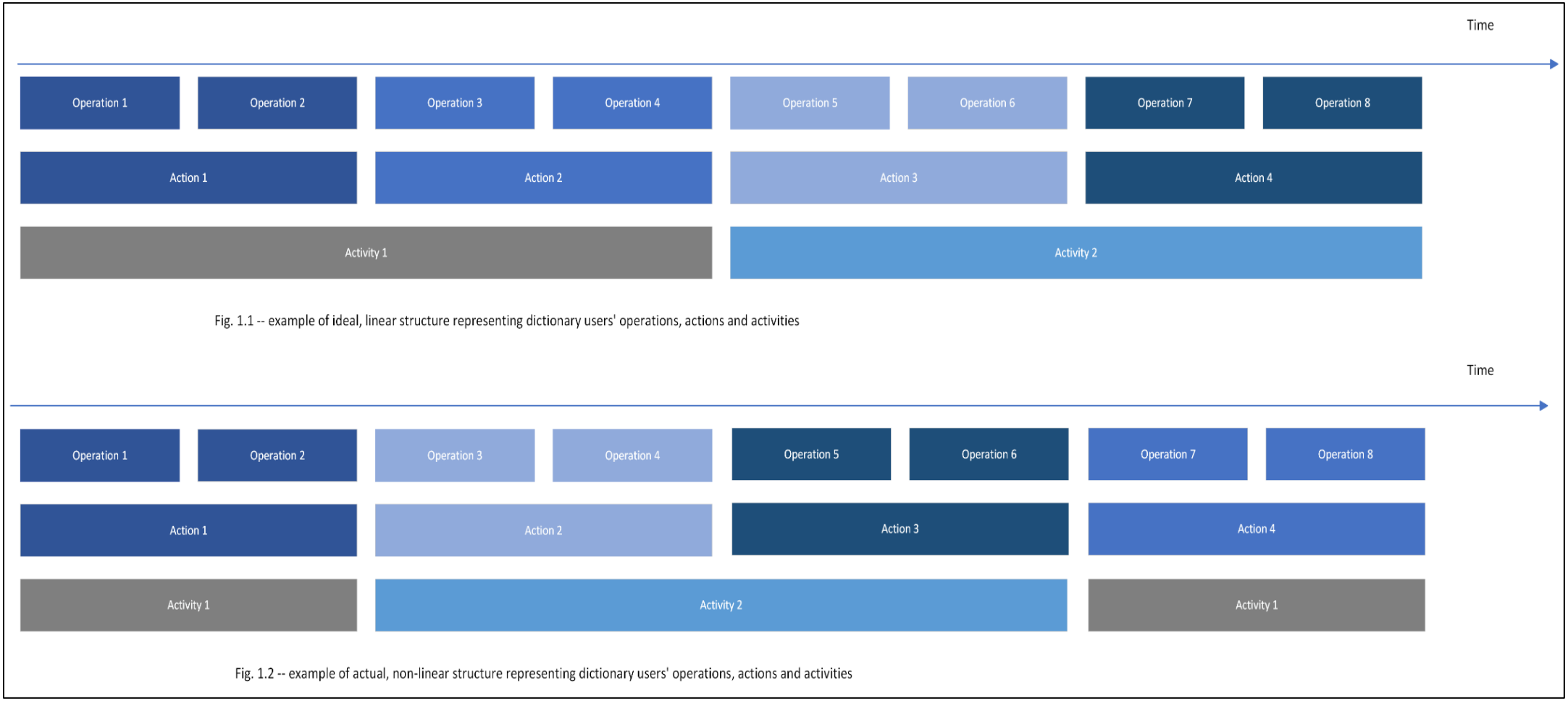

Fig. 1.1 -- example of ideal, linear structure representing dictionary users' operations, actions and activities

Fig. 1.2 -- example of actual, non-linear structure representing dictionary users' operations, actions and activities

The actual, non-linear structure of recorded data, presented in Figure 1.2, shows the benefits of using a multi-tiered annotation and transcription system. Instead of analyzing isolated sequences of operations, which might lead to erroneous conclusions concerning their purpose or justifiability, it offers a tool to detect more complex processes, no matter how dispersed they seem to be throughout the recording. In order to correctly record potential complexity of the aforementioned phenomena, we suggest using codes for operations which reflect their linear progression as well as their affiliation to a specific action and activity. Therefore, Operation 6 from Figure 1.2 would be recorded as Operation 3.3.2 (operation.action.activity), while operation 8 from the same figure would be 8.4.1. By assigning three numbers to one operation, it is possible for the researcher to filter all operations by their order of occurrence or by their membership in the class of Actions and Activities.

Success/failure attributes

In the case of studies into recorded learner lookups, success and failure can be interpreted on different levels. For instance, a successful lookup (i.e., finding the information sought in a dictionary) does not mean that the learner was looking for the **right** information to complete a given task. The multi-tiered nature of our standard offers the opportunity to replicate the complexity of the lookup process. Success/failure values can be assigned to all three levels of transcription:

- **Operations** – since operations are unconscious, normally they are considered to be neutral. However, there are some cases in which it might be justified to assign the value of "failure" to them. Some of them might be related to:
    - technical problems – for instance, when one clicks on a link, and the website crashes,
    - wrong manual execution of a given task – for instance, when one inadvertently clicks on a commercial rather than the intended button,
    - misspellings[1].
- **Actions** – successful actions are sets of operations which show that the learners achieved their intended goal. In this case, the researcher should not focus on the general objective, but on the short-term goals. The aforementioned description of a learner who conducts a successful dictionary search for a given word, regardless of whether the word itself is relevant in the context of the task, provides a clear example of a successful action (and potentially unsuccessful activity). On the other hand, the action should be considered unsuccessful if the learners do not find the information sought – either because it is not in a given dictionary or because they failed to locate the information on the webpage.
- **Activities** – in successful activities, learners produce correct language (in the case of production tasks) or gain the understanding of the previously-unknown language material (in the case of receptive tasks). This category relates to the ability to use the information found in electronic sources in order to complete a given task. Since a successful activity might comprise some unsuccessful actions or operations, the researcher gains the ability to analyze the factors that determine the overall outcome of a given look-up. Examples of such factors include, but are not limited to:
    - the number of operations per activity,
    - the number of actions per activity,
    - the ratio of successful to unsuccessful actions/operations,
    - the average duration of an action/operation within a given activity.

[1] It may be unclear whether misspelling is always a failed operation or a separate type of failure that may but does not have to most commonly occur at the level of operations. We included it here, but it could also be assigned a separate label.

We propose that the Succes/Failure value be assigned to actions and activities, with an option to include operations if deemed necessary by the researcher. We theorize that there might occur cases in which neither value can be attributed to these categories; such instances might require the third option which would express uncertainty as for the assessment of a given phenomenon. Such marker might be general ("don't know") or more specific ("insufficient data", "confusing results" etc.). However, since we have not encountered similar problems yet, the inclusion of the third category remains a theoretical possibility.

Sources of lexico-grammatical information

One of the most important pieces of information in dictionary-use studies is the choice of sources used by the subjects. The following labels evolved as a result of the standard application in practice.

Name	Definition and/or example	Remarks	C/O
Name of the source	Name or website address of a given source of lexico-grammatical information.	This label might be assigned to actions or operations. The level of operations, though more time-consuming to record, might be beneficial if the learners use a tool within a tool (e.g., a spell-checker is activated in a dictionary search box).	C
Dictionary?	In our research, we focused on the proportion of dictionary to non-dictionary sources in learners' lookups. Non-dictionary sources are materials which were not designed as dictionaries. They include search engines, message boards, videos, etc.	Binary label: Yes/No	O
Search-engine assisted?	In some cases, learners know from the beginning which source they intend to use, and even if they use a search engine (Google, Bing etc.), they only do it in order to look up the source. Another strategy is to "google" a given word/phrase and click on a dictionary/source suggested by the search engine. *Search-engine assisted* refers to the latter case, so the label does not refer to the use of a search engine per se; instead, it shows what it is used for – either to look up the name of a dictionary or the term sought itself.	This binary label ought to be assigned to actions. All operations carried out within a dictionary which was suggested by the search engine need to be considered from the point of view of potential unfamiliarity of a learner with the aforementioned tool.	O

Table 2. Labels for categories related to sources of lexico-grammatical information

While the list of the optional labels presented in the table might be extended and modified, the *Name of the source* label remains a crucial source of basic information about learner lookup activities. The number of operations performed within a specific source might later be correlated with other factors, such as success rate or average duration of a single operation.

Special labels

Our standard introduces a high degree of flexibility as regards to the phenomena that a researcher aims to investigate. The structured multi-tiered transcription standard allows one to introduce any number of labels and determine whether a given phenomenon occurs at the level of operations, actions, or activities.

Table 3 depicts labels used in our research which were mostly concerned with dictionary skills of EFL learners. They are an example of customized categories that still fall within the scope of the proposed framework.

Name	Definition and/or example	Remarks	C/O
Failure to spot relevant information	A learner opens a webpage with a correct answer, but they seem not to see it. Instead, they keep looking for the answer in other sources.	Label assigned to actions; analysis of preceding and following actions is necessary to confirm the occurrence of this phenomenon.	O
Incorrect L1 in bilingual sources	A learner uses incorrect L1 forms while typing a query in an online consultation form. The issue might be spelling, or word choice.	Normally assigned to operations.	O
Definition-based lookup	Learners use a search engine to type a definition of a word and see if any relevant pieces of information are shown in search results.	Label assigned to actions	O
Table 3. Optional, study-specific labels			

Each label presented in Table 3 can be analyzed in relation to operations, actions, and activities. For instance, if a researcher suspects that *failure to spot relevant information* occurs due to learner fatigue, they may consider the following questions:

1. **When** does the phenomenon occur during the task?
    a. How many operations/actions/activities were carried out (on average) before it first appeared?
    b. Is value from point a. similar for all the subjects, or are there any significant differences?
    c. Are there any signs of learner fatigue just before the occurrence of this phenomenon? The increase in the number of operations per minute or the constant increase in the number of seconds necessary to carry out an operation might be good indicators of this phenomenon.
2. Does the learner use the source in which the phenomenon was observed for the first time? Do they generally have their preferred sources, or do they prefer to click on the first link suggested by the search engine? What percentage of lookup actions involve sources selected by the search engine and what percentage includes other strategies?

Data recording and structuring

While we envisage that in the future AI assistants coupled with log files might simplify the process of data collection, as of 2023 it seems that manual recording is the most viable and reliable option. The data might be recorded on a simple spreadsheet, but other pieces of software can be used as well. If one wants to use a video annotation tool, such as Elan (https://archive.mpi.nl/tla/elan), we suggest

that each label be assigned to a different annotation tier. Such a layout seems to be intuitive for both the annotators and the researchers. An example of this approach is presented in Figure 1.1 and 1.2 which illustrates the way of annotating the timeline of the recording for operations, actions, and activities.

Conclusions

Our standard provides a versatile tool for researchers interested in using recordings of on-screen activity depicting dictionary use for the purpose of studies in pedagogical lexicography. The system has been perfected by us during our research into learner dictionary skills. The major benefits include:

1. The ability to combine high-granularity and low-granularity perspectives on the data and to find relations between the two.
2. The flexibility to design customized labels which produce data that can be easily analyzed from the two aforementioned perspectives.
3. The insistence on analyzing all the learners' decisions and their execution in context, which significantly reduces the risk of misinterpreting the results.
4. The transferability and comparability of the core data across different studies which might deal with vastly different phenomena. This results in the high potential for reusability of data sets for other studies.
5. The ease of access to output data for people outside the field of pedagogical lexicography by conforming to other well-established sets of standards (cf. Geisler & Slattery, 2007).

Sources

Geisler, C. & Slattery, S. (2007). Capturing the activity of digital writing: Using, analyzing, and supplementing video screen capture. In H. A. McKee & D. N. DeVoss (Eds.), *Digital writing research: Technologies, methodologies, and ethical issues* (pp. 185–200). Cresskill, NJ: Hampton Press.

Acknowledgement of the funding source

This research is funded by National Science Center, Poland. Research grant:
Miniatura-6 2022/06/X/HS2/00547

Narodowe Centrum Nauki – Miniatura-6 2022/06/X/HS2/00547