Konin Language Studies
Faculty of Humanities and Social Sciences, University of Applied Sciences in Konin, Poland
KSJ 12 (1). 2024. 55-72
http://ksj.konin.edu.pl
doi: 10.30438/ksj.2024.12.1.3

Using an Activity Theory-based framework to standardize annotation and transcription of language learners' online lookups

Marek Molenda ✉
University of Lodz
https://orcid.org/0000-0001-5468-1524
marek.molenda@filologia.uni.lodz.pl

Anastasiia Anisimova
University of Lodz
https://orcid.org/0000-0001-7019-4697
anastasiia.anisimova@edu.uni.lodz.pl

Abstract

In this paper the authors present their Activity Theory-based framework for annotating and transcribing language learners' data recorded by means of screen casting during studies conducted in the field of pedagogical lexicography. The benefits of the framework are presented in the context of improving customizability and comparability across various studies which use screen casting to provide an account of learner lookup behavior. The authors present selected examples from their own research conducted with the use of the proposed framework in order to illustrate how a three-tiered approach comprising operations, actions and activities can be used to record lower-order as well as highly complex phenomena. Selected pedagogical implications included in the final section showcase possible applications of the AT-based framework for developing language learners' dictionary skills.

Keywords: pedagogical lexicography; EFL; digital dictionary

1. Introduction

Observation-based studies which aim to record language learners' interactions with a dictionary have long been recognized as one of the most valuable methods in pedagogical lexicography. In 1997, Zikmund described their potential for providing unbiased research data in the following way:

> The major advantage of observation studies over surveys, which obtain self-reported data from respondents, is that the data do not have distortions, inaccuracies, or other response biases due to memory error, social desirability, and so on. The data are recorded when the actual behaviour takes place. (p. 265)

Nesi (2000, p. 33) added that this approach is particularly useful for "generating hypotheses" concerning the use of dictionaries by language learners.

Before the advent of affordable and versatile computer technologies, observation-based studies were considered to be problematic due to technical constraints. Hatherall (1984) described these limitations in the following way:

> Ideally [...] the researcher would actually watch users in action. But this, too, causes problems. Under such conditions it would probably be difficult for the subjects to behave normally as users. Also, it is unlikely that all the information the researcher needs would be retrievable via the visual medium. And finally, such an exercise is so time-consuming that the sample is likely to remain unrepresentatively small. (p. 184)

One of contemporary data recording techniques, namely screencasting, seems to address these concerns exceptionally well. The process of recording learners' screen while they perform lookups provides rich context for every decision made by the subject. Coupled with eye-tracking or log files, it has the potential to offer a comprehensive account of one's dictionary experience.

However, it seems that pedagogical lexicography lacks a unified framework for annotating data recorded in the form of screencasts. In this paper, we would like to postulate that the use of a structured annotation framework based on Activity Theory (AT) might help make research results more accessible while providing researchers with more opportunities to conduct their own analysis of the existing datasets.

2. Activity Theory for screen casting

While we have not, as of 2023, identified any research projects (with the exception of our study) that use Activity Theory-based framework for annotating screen

casting data obtained from learners using dictionaries, the idea itself is not new. Activity Theory has been one of the major tools for annotating and processing screen casting data since the 1990s. In 1996, Kuutti proposed that AT be used as a standard for research into Human-Computer Interaction due to its ability to accurately record low-level as well as higher-order phenomena in the context (social, material, etc.) in which they occur. In the same year, Bellamy claimed that AT should be used in designing digital tools and activities for learners of all subjects.

The idea of combining AT with screen casting technology in order to provide an account of actual human behavior when faced with a task in a digital environment was described by Bødker (1996). The author combined ethnographic data about the interactions of office workers with an AT-based framework to map operations performed in a word processor window in order to provide an account of the use of digital office tools. While similar AT-basesd screen casting studies were later conducted in the field of language and general education, they focused mostly on the process writing (e.g., Kessler, 2020, focused on L2 writing strategies) or group project work, with dictionaries occasionally mentioned as one of the tools.

Therefore, the standards for combining AT and screen casting for recording Human-Computer Interaction were largely formulated outside of the FL/L2 educational context. The version used in our research is based upon a paper published by Geisler and Slattery (2007), in which the authors list the features of AT which make it conducive for recording and interpreting the data gathered by means of screen casting. This list was modified in order to adapt its features to the needs of pedagogical lexicography (Molenda & Anisimova, 2023):

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 (Nesi, 1999) 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.

The major advantages of this framework in the context of observation-oriented research projects conducted in the field of pedagogical lexicography are its adaptability and transferability. The former refers to the fact that it allows the researchers to focus on various phenomena which they wish to investigate. A number of individualized labels can be attached to sets of operations, actions and activities, some of which are presented in the following sections of this paper. On the other hand, the imperative to record the events visible on the screen by means of a standardized three-tiered annotation structure makes it possible to approach data from different perspectives or easily compare results from various independent research projects. In other words, this approach allows one to create standardized datasets which – while losing none of their original research angle – are easily comparable with similar studies conducted by other researchers.

The need to maintain a standardized procedure for recording language learners' operations, actions and activities seems to be especially important in the context of an increase in the interest in screen casting tools observed in pedagogical lexicography. Upon examining recent publications, one may conclude that the research frameworks used in them produce data which is difficult to compare due to different coding conventions. These discrepancies might be demonstrated by the example of authors' approach to the structure of lookups and their division into activities, actions and operations. The differences are presented in Table 1 which uses the AT-based framework as a reference point.

Table 1 Terminology used to describe activities, actions and operations in recent studies

Source	Approach to activities	Approach to actions	Approach to operations
Bailey & Withers, 2018	No specific name given.	Referred to as "interactions" (p. 181) or "uses/use" (p. 182-183).	The authors avoid any labels, so all operations, e.g., "use of a synonym finder" or "typing" are referred to as "codes" (p. 179) in the context of the study, and "uses/use" if they are discussed in the broader context. It seems that the lines between actions and operations are blurred.
Müller-Spitzer et al., 2018	Activities are commented upon, but they are not given any specific names.	Referred to as "queries" or "searchers" or "search actions". These terms seem to be used interchangeably (p. 292, 298).	Referred to as "actions" in the context of think-aloud protocols (p. 295) and "trials" if they refer to sentence edits (p. 298).
Gilquin & Laporte, 2021	Referred to as "search" when it requires the use of "multiple tools" (p. 6). The name "Activity" is used to describe the total time spent on writing, looking for information or performing other tasks.	Referred to as "search" when it describes one lookup with a "one tool" (p. 6).	Referred to as "queries"; a non-query type of an operation was not given a specific name.
Chen & Liu, 2022	Referred to as "dictionary lookup behavior" (p. 483).	Referred to as "occasions of use" (p. 483), "searches" (p. 479) or "search patterns (p. 484) if they were assessed in terms of successful completion.	Referred to as "moves" (p. 475).

As shown in Table 1, the lack of a unified set of standards might lead researchers to some confusion as to how to interpret the results of each study. One of the best examples is the term "search" which seems to serve a slightly different function in each study in which it appears. Of course, we do not postulate that the researchers resort solely to the three AT-based categories, i.e., activity, action and operation. It would be highly impractical to refrain from using other terms, such as "search", "look-up" or "query". However, we recommend that all of these "traditional" terms be precisely defined in relation to the AT framework in order to keep our work consistent and comparable.

3. AT-based framework applied in practice – the study

Having listed the benefits of the AT-based framework on a macro level, we would like to present its application in an actual study. The project was designed to generate hypotheses regarding language learners' use of online dictionaries and other consultation tools. Therefore, research questions were relatively broad; the aim was to (1) obtain information about the tools used by the learners (2)

record successful and unsuccessful actions and describe what could have contributed to the failure/success in a given case (3) record other phenomena which could be potentially interesting in the context of learners' dictionary skills (Nesi, 1999). The following description, however, does not report the full results of our study. Instead, it aims to present selected cases in which the AT-based framework might be especially useful for a researcher.

Table 2 Examples of operations, actions and activities

Name	Definition and/or example	Remarks
Operation	Single event, e.g., *click on the "search" button*; *type the word "guard" in the textbox*	Every operation had 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.
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 dictionary* *type "caterpillar"* *click on the search button* *analyze the entry* *scroll down the entry* *scroll up* *close the window*	An action might span across various sources provided that the aim remains unchanged.
Activity	Sets of actions carried out in order to solve a given problem (i.e., to fill the 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 fully explained (e.g., as looking for new information, verification, etc.) solely in the context of activities which outline the global goal.

Our subjects were eight female English Philology students at the university of Łódź, with their attested CEFR language level at C1/C1+. All of them completed a relatively challenging online task which in its form closely resembled their PNJA (Practical English) online exam. The task comprised the following components: keyword transformations (five sentences), a missing word activity (five gaps) and error correction task (five sentences). The subjects performed the tasks in a classroom using the same hardware that they would normally use during their PNJA exams. All of them were informed that they could use any

online consultation source – dictionary and beyond – that they found suitable. In a follow-up questionnaire, 100% of the respondents confirmed that online dictionaries were their major sources of lexical information, which made their task and its context relatively close to real-life conditions postulated by Hatherall (1984). The participants were also informed that their on-screen activity would be recorded for research purposes and that they have the right to preview their recordings and have them permanently deleted upon request.

The recordings were annotated with Elan software (2023) and exported to spreadsheets for further analysis. Three basic Elan annotation tiers were operations, actions and activities. Examples of these categories are presented in Table 2.

4. Non-linear data structure

It was discovered that one of the major advantages of the aforementioned three-tiered structure is the fact that it allows one to provide an account of seemingly chaotic behavior observed in the learners. While in an ideal use case, dictionary users would always finish one action before proceeding to the next one, in reality they oftentimes act more chaotically, which leads to false-starts, re-starts, interrupted lookups, etc. Figure 1.1 presents an optimal scenario, in which operations would always form consecutive actions, and actions would always combine to form activities in a linear way without any overlaps. Figure 1.2, on the other hand, shows the structure of actual data, in which it was not uncommon for the subjects to interrupt a given action or activity in order to return to it later (1.2).

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 used 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 6.3.2 (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 the recorded operations by their order of occurrence or by their membership in the class of actions and/or activities.

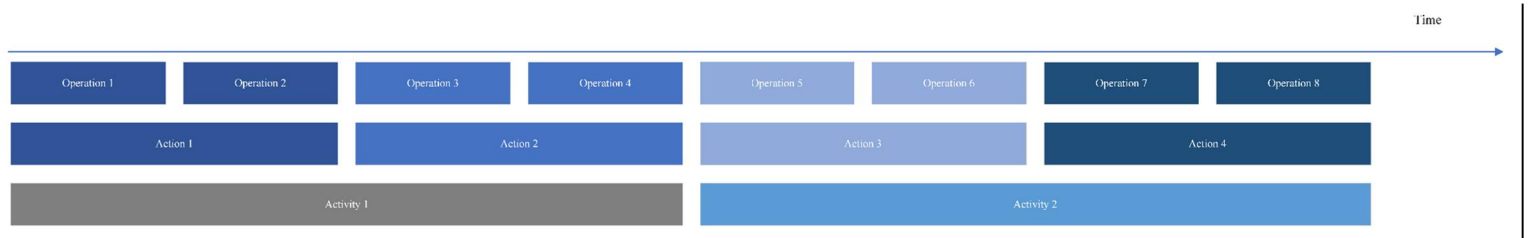

Fig. 1.1 -- example of ideal, linear structure representing dictionary users' operations, actions and activities

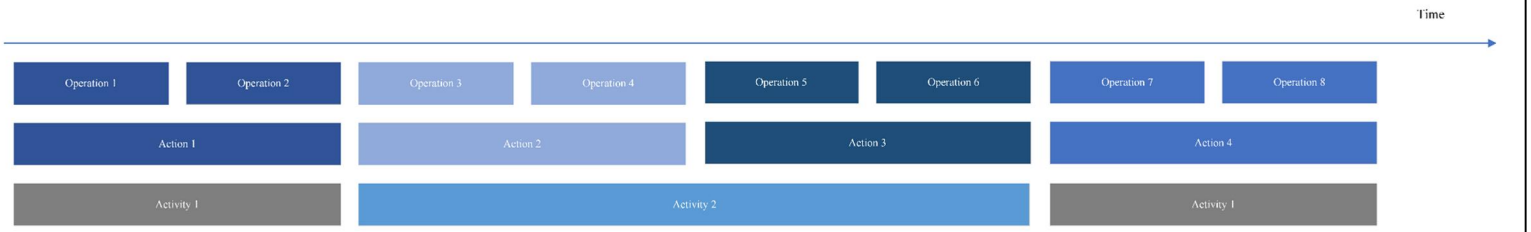

Fig. 1.2 -- example of actual, non-linear structure representing dictionary users' operations, actions and activities

Figure 1.1 and Figure 1.2 An optimal scenario vis-à-vis a sample of an actual data structure

5. The problematic notions of "failure" and "success"

In the case of studies into recorded learner lookups, success and failure can be interpreted on different levels. For instance, a successful lookup, such as 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 the AT-based standard offers the opportunity to account for the complexity of the lookup process. Success/failure values can thus 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.
- 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 situation when a learner 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 a 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 lookup. 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.

The examples presented in this section depict various levels of "success" and "failure" which can be explained by the fact that they refer to different dictionary skills (Nesi, 1999). Thus, we postulate that instead of using simple categories, such as successful or unsuccessful searches (cf. Chen & Liu, 2022), the researcher should build learner profiles which take into account students' key strengths and weaknesses (Lew, 2021).

6. Choice of sources

The three-tiered structure comprising operations, actions and activities provides a more in-depth insight into the choice of sources. Analyzing sequences of operations that lead to accessing a given source may show one how and to what extent the use of a particular website or tool was a conscious choice and to what extent it depended on the circumstances beyond subjects' control. In the former case, the learner either opens a given tool directly or uses a search engine to find it. This set of operations suggests that they know exactly which tool they are planning to use to accomplish a given task. On the other hand, if the learners use a search engine to look up phrases such as "[word/phrase sought] dictionary" (e.g., "cone off dictionary") and then they immediately click on the first result, it is the search algorithm that determines the outcome for them.

Therefore, it seems insufficient to ask about the choice of sources used by the subjects; on the contrary, it is necessary to determine which sources were selected as result of a conscious decision and which were accessed with the help of the search algorithm. This binary parameter, which we named "search engine-aided" (or "Google-aided", as Google was the only search engine used by our subjects), may have an impact on how successful given lookup activities or actions are. In addition, using a search engine-aided lookup as the only strategy might suggest subjects' deficiencies as regards one of key dictionary skills, namely Skill 6 from sub-set 2.2 from Nesi's classification -- "Deciding which dictionary is most likely to satisfy the purpose of the consultation" (cited in Lew, 2013, p. 18).

7. Developing study-specific labels

Apart from the aforementioned basic data structure, the AT-based framework makes it possible to define additional categories which represent other phenomena observed during the study. These categories, also referred to as "labels", should overlap with actions and activities if possible. In such a case, it may be said that they "are assigned to" or "are attributes of" a specific action/activity.

If the labels do not meet this criterion, they can still function successfully as separate entities. The only truly indispensable requirement is that they should always comprise operations. Owing to this solution, the labels remain easy to interpret by other researchers while retaining their original character, which translates to comparability and customizability across research projects.

Examples of labels which were defined in the course of our study are presented in Table 3:

Table 3 Selected labels form our study

Name	Definition and/or example	Alignment
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.
Incorrect L1 in bilingual sources	A learner uses incorrect L1 forms while typing a query in an online consultation tool. The issue might be related spelling, word choice or word form.	Normally assigned to operations.
Definition-based lookup	Learners use a search engine to type a definition of a word/phrase and check if any relevant pieces of information are shown in search results.	Label assigned to actions

8. Template for a data table

Following Geisler and Slattery's example (2007), we would like to present a simplified template for a data table which was used in the research, along with a sample of descriptions of actual activities, actions and operations recorded during the study.

The core part of the data table are the smallest indivisible AT units, i.e., operations. They are described in detail by the researcher in the "Operation" column in order to facilitate access to the specifics of user behavior at a given moment. Additional context is provided by the "Artifact", "Task ID" and "Relevant information shown?" columns. This last column features mostly blank cells – it is only completed at the level of individual operations when the subject starts browsing the list of results. This is done in order to determine whether the information sought was actually available to the learner for the duration of a given operation.

Other columns with binary Y/N (Yes/No) values are related to the categories described in the previous section, and they include (un)successful activities and actions as well as the information on whether a given lookup was aided by a search engine. The data is recorded for every row to facilitate information retrieval, should the table be converted to an SQL database.

Table 4 Simplified version of three excerpts from authors' data table

Subject ID	Time stamp	Activity ID	Activity description	Action ID	Action description	Operation ID	Operation	Task ID	Relevant info	Artifact	Activity successful?	Action successful?	Action aided by a search engine?	Additional labels
2	13:32	3	Looking for the missing word	3	Looking up "loath" to find its collocates	8	Type "loathe" in MED search box and execute search	1.4		MED	Y	Y	N	
2	13:38	3	Looking for the missing word	3	Looking up "loath" to find its collocates	9	Browse the entry	1.4	Y	MED	Y	Y	N	
2	13:40	3	Looking for the missing word	3	Looking up "loath" to find its collocates	10	Open the tab with a task	1.4		Task	Y	Y	N	
2	13:44	3	Looking for the missing word	3	Looking up "loath" to find its collocates	11	Open the tab with MED entry for "loathe"	1.4		MED	Y	Y	N	
2	13:48	3	Looking for the missing word	3	Looking up "loath" to find its collocates	12	Open the tab with the task and type "to" in 1.4	1.4		Task	Y	Y	N	
[...]														
2	15:33	4	Looking for the missing word	7	Looking up the entire gapped sentence	22	Open the tab with the task and copy gapped sentence 1.5	1.5		Task	Y	N	Y	5
2	15:36	4	Looking for the missing word	7	Looking up the entire gapped sentence	23	Paste the sentence into Google engine	1.5		Google	Y	N	Y	5
2	15:43	4	Looking for the missing word	7	Looking up the entire gapped sentence	24	Browse the list of results	1.5	N	Google	Y	N	Y	5
2	15:53	4	Looking for the missing word	7	Looking up the entire gapped sentence	25	Click on Google search box to expand the list of suggestions	1.5	N	Google	Y	N	Y	5
2	15:55	4	Looking for the missing word	7	Looking up the entire gapped sentence	26	Continue browsing the list of results	1.5	N	Google	Y	N	Y	5
2	16:00	4	Looking for the missing word	7	Looking up the entire gapped sentence	27	Open the tab with the task and move cursor	1.5		Task	Y	N	Y	5
[...]														
2	19:53	4	Looking for the missing word	12	Browsing practice set to confirm the answer	48	Use CTRL + F to look up the word "store" in the set of flashcards	1.5	Y	Quizlet	Y	Y	N	
2	20:14	4	Looking for the missing word	12	Browsing practice set to confirm the answer	49	Go back to the task and leave the answer unchanged	1.5		Quizlet	Y	Y	N	

Other columns with binary Y/N (Yes/No) values are related to the categories described in the previous section, and they include (un)successful activities and actions as well as the information on whether a given lookup was aided by a search engine. The data is recorded for every row to facilitate information retrieval, should the table be converted to an SQL database.

Time stamps used in the table reflect the minute and second of the recording in which a given operation was initiated. The duration can therefore be calculated by subtracting its time stamp from the value for the preceding operation. For instance, Operation 24 in Table 4 lasts 10 seconds (from 15:43 to 15:53). The subtraction may be performed automatically in a spreadsheet and added as another column to the database or it can be recorded at the moment of transcribing the data if one uses time-aligned annotation software, such as ELAN.

The column named "Additional labels" contains codes for different patterns which were deemed worthy of recording by the researchers. In Table 4, label "5" is the code for "cheaping" – a term coined by Jan Volín (personal communication, November 16, 2016) to describe a situation in which a learner faced with a vocabulary-oriented task decides not to look for the right word/structure, but instead they paste the entire sentence into a search engine in hope of finding the answer key. This and other phenomena were systematized and encoded by us in this project, but "Additional labels" might be replaced or complimented with a column for text comments if necessary.

The picture that emerges from three excerpts presented in Table 4 shows a subject who firstly performs a lookup in the online version of Macmillan English Dictionary (MED), then relatively easily finds the right collocation for the verb "loathe" and copies the correct answer to the digital answer sheet. Approximately two minutes later, the learner recourses to cheaping, and they seem to study not only the page with results displayed by Google search engine, but also the drop-down list with search suggestions and related keywords. This particular action does not provide relevant information, but the data indicates that it is a part of an activity which finally lead to the successful completion of the task. Upon completing Operation 27, the learner shifts their attention to other tasks, but they return to the fifth sentence of the first exercise (Task ID 1.5) in Operation 48. By then, they had already located a Quizlet set with vocabulary practice for advanced learners (operations omitted in Table 4), and since it proved to be useful for another task, a decision is made to browse it with a simple page content search. This strategy finally leads to the successful completion of the task.

As shown in this example, AT-based framework made it possible to isolate an activity which was dispersed throughout the timeline. However, analyzing it without rich contextual information provided in the table would be a mistake, since it was precisely the impact of artifacts discovered while performing other tasks that made its completion possible.

9. Selected findings

This section presents the application of our framework in practice. It does not aim to show comprehensive research results, but rather to highlight the usefulness of the AT-based framework in the context of the aforementioned research questions.

Q1: The choice of tools

The AT-based framework makes it possible to obtain data that is more in-depth than a list of all the tools used. By counting the number of operations per tool we were able to determine the actual usage. The results are presented in Figure 2. This data could be augmented with additional temporal measures, i.e., the number of minutes and seconds spent using a given tool.

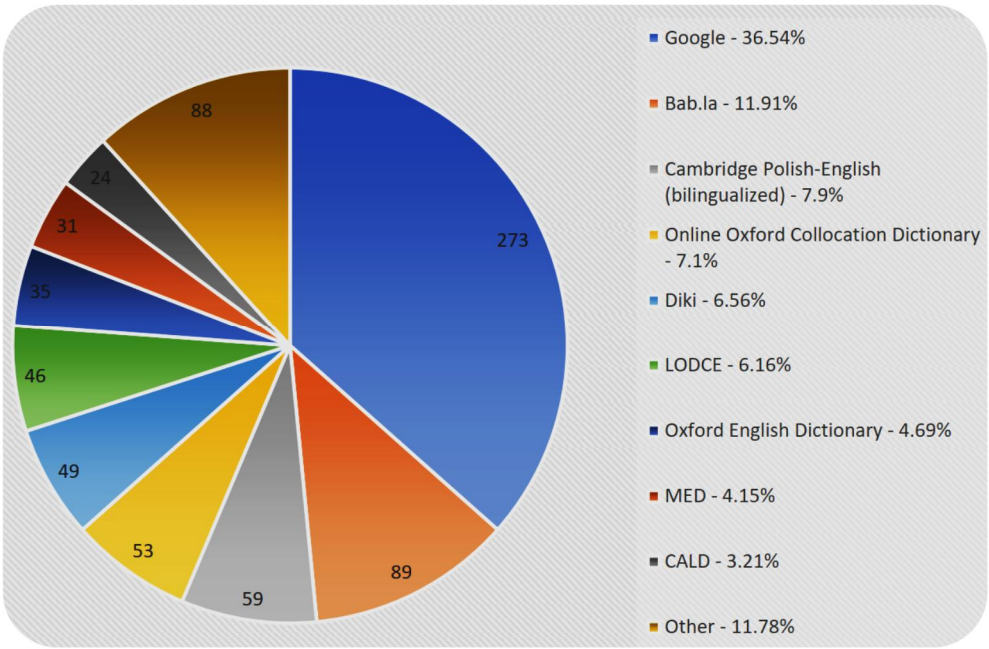

Figure 2 Number of operations recorded for 10 most extensively used tools

The data presented in Figure 2 suggests that the search engine (Google) accounted for slightly over 1/3 of the lookup operations. However, its impact seems to be underestimated if one was to assess the usage by this metric alone. In order to better understand Google's impact on our subjects, we decided to compare the number of operations in search engine-aided lookup actions to the independent ones. The results indicate that 48% (395 operations) were aided by Google, while the other 52% (421) remained independent.

Q2: Successful and unsuccessful actions

In order to record and assess the effectiveness of actions, it is necessary to rely on the two aforementioned features of the AT-based framework. Firstly, it needs to be accepted that actions might be interrupted, so their recommencements (if they exist) need to be correctly identified. Secondly, one ought to be able to separate the assessment of actions from activities and operations. In the case of our study, 54.13% of activities were assessed as successful. This value might be further compared against success scores for activities in order to assess whether the unsuccessful actions are a natural part of the lookup process or whether their accumulation tends to lead to unsuccessful activities.

Q3: Recording other interesting phenomena

As was stressed in the previous sections, labels for interesting phenomena may, but do not have to be aligned with actions. An example of a phenomenon that comprises fewer operations than the action which it belongs to is *switching*. This name refers to the situation in which the subject seemingly mindlessly switches between the browser tabs or windows without taking enough time to analyze the information that they contain (in our study, learners would typically spend 0.5-1.5 seconds per tab/window). While the exact nature of this phenomenon is unclear, it might be hypothesized that it might have been caused by cognitive overload.

By contrast, some other phenomena observed were well-aligned with actions. One of them was an interesting strategy of a definition-based search performed by one learner. Instead of looking up words, they decided to use a search engine to type the definition and explore the results. We believe that this strategy might be used much more often with AI-based tools as they continue to gain popularity among learners and students.

10. Summary and pedagogical applications

In this paper we attempted to demonstrate that Activity Theory-based annotation and transcription framework for data recorded by means of screen casting is a beneficial solution for studies in the field of pedagogical lexicography. We started by exploring benefits on the macro scale, such as the enhancement in comparability and transferability of various research projects. Then, we proposed the basic rules for such annotation (for more specific guidance, see Molenda & Anisimova, 2023). Finally, we presented some research examples obtained by this method in order to demonstrate its usefulness in practice.

Following Hatherall's postulate (1984), we believe that research into actual learner behavior is the most valuable approach when one wants to use the results obtained

in order to propose sound solutions to learners' problems. For instance, the data presented in this study might lead the teachers to the following conclusions:

1. The learners should be encouraged and taught to choose lexical sources more consciously. They need to be made aware of the fact that using pages suggested by the search engine algorithm means exposing oneself to a different interface multiple times, which might diminish the ability to successfully locate information on the page (cf. Table 3).
2. The phenomenon of switching necessitates more research, but if one assumes that it is related to learner fatigue, it may be suggested that the learners be trained in recognizing it as a symptom of a bigger problem which needs to be addressed – by means of physical activities, relaxation etc.
3. Dictionary skills training should include examples of successful actions and activities. For instance, the definition-based search which, to the best of our knowledge, is given little attention in language textbooks, might be one of the most beneficial dictionary skills of the future – especially in the context of chatting with Artificial Intelligence bots.

Again, we would like to stress that these results are just selected examples of implications that might be formed owing to the use of an AT-based framework. For the comprehensive list, see Molenda (2021).

11. Conclusion

We believe that the adoption of the AT-based framework for screen recording would be a step in the right direction for pedagogical lexicography. In the era of rapidly evolving consultation tools which can be used in the role of a dictionary, a unified system like this is likely to facilitate and expedite our research efforts.

12. Acknowledgement of funding source

This research was funded by National Science Center Poland. Research grant number: DEC-2022/06/X/HS2/00547.

References

Bailey, C. & Withers, J. (2018). What can screen capture reveal about students' use of software tools when undertaking a paraphrasing task?. *Journal of academic writing, 8*(2), 176-190. https://doi.org/10.18552/joaw.456

Bellamy, R. K. E. (1996). Designing educational technology: Computer-mediated change. In B. A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 123-146). The MIT Press. https://doi.org/10.7551/mitpress/2137.001.0001

Bødker, S. (1996). Applying activity theory to video analysis: How to make sense of video data in HCI. In B. A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 147-174). The MIT Press.

Chen, Y. & Liu, S. (2022). Exploring the use of an online bilingual dictionary in EFL writing. *International Journal of Lexicography, 35*, 468-490. https://doi.org/10.1093/ijl/ecac007

Elan (Version 6.7) [Computer software]. (2023). Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Hatherall, G. (1984). Studying dictionary use: Some findings and proposals. In R. R. K. Hartmann, (Ed.), *LEXeter '83 proceedings* (pp. 183-189). Niemeyer.

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). Hampton Press.

Gilquin, G. & Laporte, S. (2021). The use of online writing tools by learners of English: Evidence from a process corpus. *International Journal of Lexicography, 34*(4), 472-492. https://doi.org/10.1093/ijl/ecab012

Kessler, M. (2020). Technology-mediated writing: Exploring incoming graduate students' L2 writing strategies with activity theory. *Computers and composition, 55*. https://doi.org/10.1016/j.compcom.2020.102542

Kuutti, K. (1996). Activity theory as a potential framework for human-computer interaction research. In B. A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 17-44). The MIT Press. https://doi.org/10.7551/mitpress/2137.001.0001

Lew, R. (2013). Online dictionary skills. In Kosem, I. et al. (Eds.), *Proceedings of eLex 2013* (pp. 16-31).

Lew, R. (2021). *Recenzja rozprawy doktorskiej mgra Marka Molendy*. University of Lodz Repository.

Molenda, M. (2021). *Investigating the process of EFL lexico-grammatical consultation in the Web 2.0 environment*. Doctoral thesis. University of Lodz Repository. http://hdl.handle.net/11089/40439

Molenda, M. & Anisimova, A. (2023). *Annotation and transcription standards for research into the process of EFL lexico-grammatical consultation in the Web 2.0 environment*. University of Lodz Repository. http://hdl.handle.net/11089/48096

Müller-Spitzer, C., Nied Curcio, M., Domínguez Vázquez, M. J., Silva Dias, I. M., & Wolfer, S. (2018). Correct hypotheses and careful reading are essential: Results of an observational study on learners using online language resources. *Lexicos, 28*, 287-315. https://doi.org/10.5788/28-1-1466

Nesi, H. (1999). The specification of dictionary reference skills in higher education. In R. R. K. Hartmann (Ed.), *Dictionaries in language learning. Recommendations, national reports and thematic reports from the Thematic Network Project in the area of languages, sub-project 9: Dictionaries* (pp. 53-67). Freie Universität Berlin.

Zikmund, William G. (1997). *Business research methods*. Dryden.