

Combining Quantitative and Qualitative Methods in a Study on Dictionary Use

*Sascha Wolfer¹, Martina Nied Curcio², Idalete Maria Silva Dias³,
Carolin Müller-Spitzer¹, María José Domínguez Vázquez⁴*

¹*Institut für Deutsche Sprache*, ²*Universita degli Studie Roma Tre*, ³*Universidade do Minho Braga*,

⁴*Universidade de Santiago de Compostela*

*E-mail: wolfer@ids-mannheim.de, martina.nied@uniroma3.it, idalete@ilch.uminho.pt,
mueller-spitzer@ids-mannheim.de, majo.dominguez@usc.es*

Abstract

Many studies on dictionary use presuppose that users do indeed consult lexicographic resources. However, little is known about what users actually do when they try to solve language problems on their own. We present an observation study where learners of German were allowed to browse the web freely while correcting erroneous German sentences. In this paper, we are focusing on the multi-methodological approach of the study, especially the interplay between quantitative and qualitative approaches. In one example study, we will show how the analysis of verbal protocols, the correction task and the screen recordings can reveal the effects of intuition, language (learning) awareness, and determination on the accuracy of the corrections. In another example study, we will show how preconceived hypotheses about the problem at hand might hinder participants from arriving at the correct solution.

Keywords: research into dictionary use, observation study, language learners, quantitative and qualitative methods, online lexicographic resources

1 Introduction

In the past two decades, more and more studies on dictionary use have been published. Most of them have investigated what users appreciate about dictionaries, which dictionaries they use, which information they need in specific situations and whether relevant information can be accessed easily and quickly within the dictionary. The lexicographic community benefited considerably from these studies (Dziemianko 2014, Lew 2015a, 2015b, Müller-Spitzer 2014a). However, most research conducted so far presupposes that users indeed do consult lexicographic resources. In contrast, language teachers and lecturers of linguistics often have the impression that students use too few high-quality dictionaries in their everyday work. As such, a lot of studies on dictionary use might start at a point that many students may never reach when dealing with everyday language problems.

Against this background, we started an international cooperation project to collect empirical data about what students (starting with students of German who are native speakers of a Romance language) actually do when they correct language problems in their second language. With this study, we want to complement results from works on a larger scale (i.e. studies reaching much more participants but collecting less detailed data), like questionnaire studies (cf. Levy/Steel 2015, Müller-Spitzer 2014b). To do so, we carried out an observation study with learners of German who are native speakers of Romance languages combining screen recordings (to observe what participants do) and verbal protocols (to get an idea of the intuitions and motivations of participants) during a correction task.

In the present paper, we want to emphasize the multi-methodological approach we took while designing the study. We chose this multi-methodological approach primarily because the study is very exploratory in nature. This means that although we followed specific research questions (e.g. whether dictionaries are used at all, and if so, how they are used), we did not test pre-formulated hypotheses. Consequently, we used a relatively “free” experimental setup (no experimental factors that are being varied systematically) with a rather open task (improving L2 sentences).

In a way, this is a risky approach, because we had to make sure that the observations we make can be compared and cross-referenced with something if there are no explicit experimental conditions that suggest certain comparisons. This is where the multi-methodological approach comes into play. The empirical data measured by the different methods (correction task results, verbal protocols, and screen recordings) can be combined with and compared to each other to gain a more complete picture of the processes that might prove pivotal for a successful correction of errors in the participants’ L2.

Before we describe the combination of methods in more detail, we want to emphasize the interplay between quantitative and qualitative research methods that proved very beneficial during data analysis and interpretation. We pursue an iterative process, which might start with a more qualitative observation (e.g. the impression that some participants behave rather unsystematically when researching in online lexicographic resources) that needs to be translated into a measurable variable (e.g. the mean number of seconds a participant spends on a resource during research), a process commonly referred to as “operationalization”. When a variable is operationalized and all relevant measurements are extracted from the data, more questions might emerge (e.g., “Although participant A stays on resource B for a long time, he does not find the correct solution to the problem. Why is that?”). Now, one might find an answer in the verbal protocol of participant A, which calls again for a more qualitative, interpretative approach. Therefore, quantitative and qualitative approaches take turns, constantly leading closer to a better understanding of the representations and processes guiding research for a language correction task. As we have already pointed out, we do not think that studies presupposing the use of dictionaries are unnecessary. Quite the contrary: these studies are indeed very useful when compiling and optimizing lexicographic resources. One aim of our work was simply to put these studies into perspective in providing an impression of how relevant lexicographic resources really are when dealing with language problems in a language-learning context.

In the next section (2), we will describe the experimental setup and the methods we applied. We will also show briefly how the data was annotated and combined to allow for the multi-methodological analyses. In Section 3, we will introduce two example studies relying more on the qualitative side of the data and analyses. In Section 4, we will discuss and sum up the results presented in this paper. An article covering more quantitative aspects of the study is currently in preparation (Müller-Spitzer et al., in prep.).

2 Experimental Setup and Methods

Altogether, data from 43 participants was collected. 15 university students participated in Santiago de Compostela (Spain) and 14 people in Braga (Portugal) and Rome (Italy) respectively. All participants speak German on a CEFR level between B1 and B2. For data collection, we combined a language correction task, screen recordings of all on-screen actions, and audio recordings to prepare verbal protocols after the experiment. We handed the participants a written instruction in their native language before the experiment. Along with a detailed description of the task, the instruction contained an explicit remark that they were not graded with the study. Moreover, their

university teachers were not present in the room. We found this especially important because we wanted the participants to behave as “naturally” as possible. The instruction also contained some clues on the thinking-aloud task (see below) to make it easier for them to express their thoughts during the experiment.

One or two experimenters who did not speak the participants’ native language were present in the room at all times. One additional person who spoke the respective local language natively was also present. She or he translated questions from the participants or cues from the experimenters.

The participants worked on a standard Windows 10 desktop environment on a 15-inch notebook with German keyboard layout and a wired mouse. Google Chrome (Version 57.0) and Mozilla Firefox (Version 52.0) were available for browsing. After each participant, the browser cache and history were cleaned. We used the same notebook for all participants in each country, but set the browser language to the respective local language.

2.1 Correction Task

The main task of the participants was to correct 18 German sentences. Each of the sentences contained one error. The word(s) that constituted the error were highlighted in bold. The sentences were constructed according to the following two criteria. i) The error is typical of early learners of German with a Romance native language. ii) The error should not be easily resolvable by simply searching the web for the stimulus sentence or parts thereof. We tested this for each stimulus sentence in all participating countries beforehand.

We used a simple Excel spreadsheet containing the stimuli sentences in one column labelled “Satz” (Eng. sentence). Next to it was a blank column labelled “Korrektur” (Eng. correction), where the participants need to type their corrected sentence. For each participant, the sequence of sentences was shuffled, and we did this to avoid position effects (e.g., the first sentence always being more likely to be correct). By using standard office software instead of special experimental software, we aimed to situate the task in an environment the participants are well acquainted with. They were not allowed to use any built-in assistance software in Windows 10. We did not give the participants any time limit beforehand, but told them after 30 minutes that they had 15 minutes left to work on the task. After 45 minutes, they were told to finish the sentence they were currently working on and terminate the experiment after that.

2.2 Thinking Aloud/Verbal Protocols

While working on the corrections the participants had to “think aloud”, i.e. express their thoughts on-the-fly. This is not an easy task and some participants coped with it better than others. Whenever the participants fell silent, we gave a short cue after around 10 seconds of silence. The voice signal was captured with a high-definition external microphone. After data collection, the audio track recorded by the external microphone was spliced in as the audio track of the screen recordings. The verbal protocols were transcribed by native speakers. German translations of the verbal protocols are also available.

2.3 Screen Recordings

We used the screen recording software ActivePresenter to record all on-screen actions. We made sure that the screen recording software did not interfere with the task in any way (e.g. screen flickering or performance drops). As indicated above, the screen recordings were later synchronized with the audio recordings to allow for easier transcription and investigation.

2.4 Combining the Data

The corrections provided by the participants were annotated by two native speakers of German. Each correction was classified as “correct” (all errors have been resolved and none were introduced), “correct with errors” (all errors have been resolved but other errors have been introduced), “case of doubt” (it could not be determined without a doubt whether the answer is correct or not), “wrong” (the error was not resolved or has been replaced by another one), “not dealt with” (the participant did not attempt to correct the sentence). Initial weighted kappa (Cohen, 1968) was $\kappa = .86$, which is typically considered as very good agreement. This is also reflected by the fact that 712 of 816 cases (87.3 %) were labelled identically by the two annotators. All disagreements were resolved through discussion.

As indicated above (cf. Section 2.2), we transcribed the voice recordings of the participants. But these verbal protocols were not the only transcriptions that had to be created. To be able to combine all data sources into one dataset, we also had to transcribe the screen recordings (cf. Section 2.3). The recordings were split into discrete “actions” by two annotators who were trained in this procedure with a number of screen recordings. Due to the large amount of data (over 30 hours of video data had to be transcribed on a second-by-second level), the annotators then worked on different subsets of recordings. The smallest units of the transcribed screen recordings are single actions like opening a webpage, returning to Excel, typing a correction, entering a search string in a resource or a search engine, clicking a hyperlink within a resource or a search engine result list, and so on. Each action is associated with the timestamp in the respective screen recording. All other types of information are on a higher level than these single actions. Table 1 gives an impression of the organization of the dataset, but only lists a subset of the available columns (= variables) and rows (= actions).

Table 1: A subset of rows and columns of our dataset to illustrate the multi-level organization of data.
The column “VerbalProt” holds the verbal protocols (examples follow in Section 3).
All protocols are abbreviated in Table 1 due to space limitations.

<i>Participant</i>	<i>SentPOS</i>	<i>SentID</i>	<i>Timestamp</i>	<i>Action</i>	<i>Resource</i>	<i>VerbalProt</i>	<i>Correction</i>
B-01	1	11	41	types correction	Excel	< Text >	wrong
B-01	1	11	106	opens browser	Google	< Silence >	wrong
B-01	1	11	126	opens PONS	PONS Dictionary	< Text >	wrong
...
S-16	13	8	1537	clicks on hyperlink “sich verfahren”	PONS Dictionary	< Text >	correct with errors
S-16	13	8	1544	switches to Excel, types correction	Excel	< Text >	correct with errors

The different levels of the variables in the dataset can be seen in Table 1. Column “Participant” is identical for all actions from this participant, “SentPOS” (the position of the sentence in the Excel spreadsheet) and “SentID” (the unique sentence ID for the whole experiment) are always identical as long as the participant works on that sentence. Columns “Timestamp” and “Action” are on the lowest level. Column “Resource” holds the lexicographic resource, a search engine (all participants used Google exclusively) or Excel. If two actions are made while the same resource is open and on-screen, the entry is repeated in successive rows. Finally, column “Correction” holds the correctness annotation described above. This is identical for all entries of the respective sentence of the respective participant. As indicated above, the real dataset we are working with during the analyses holds further columns (e.g. information regarding search strings, error type information etc.) and 7,647 rows (= actions) altogether. But the subset of columns displayed in Table 1 suffices to illustrate the potential

cross-combinations for analyses. For example, one can look at the number of different resources that have been consulted during the work on sentences that have not been corrected properly. In the next step, the time that was spent on each of these resources can be examined (via the timestamps) and the associated verbal protocols can be investigated. All of this information can be extracted in an automated way, because the dataset is organized in this multi-level table format.

The organization of our dataset enables us to implement the research approach that we outlined in the introduction. Qualitative approaches alone fail if one does not know which cases are worth a closer look, or which groups of cases exist. Such identification of interesting cases or groups is a strength of the quantitative side of the approach. As soon as these interesting cases have been investigated in a more interpretative/qualitative manner, new hypotheses can be generated that can then be investigated quantitatively again.

3 Example Studies

So far, most of our remarks and explanations concerning the combination of quantitative and qualitative methods have been inevitably quite abstract, because we wanted to describe the idea of the study as a whole. In this section, we are presenting two concrete example studies to make this idea clearer. In this work we are emphasizing the qualitative research approach, but also try to show how the quantitative analyses are intertwined with the qualitative one.

3.1 Time, Language (Learning) Awareness and Determination

When looking at the data, we had the impression that students with less accurate sentences spent less time with the single resource. This gave us the idea of relating the time spent on a resource with sentence correction accuracy. We found out that the average time spent on resources plays a decisive role in the correctness of the final sentence (cf. Figure 1). With 2.4 seconds, the mean difference is quite small. Note that this difference means that – on average – the time spent on each single resource is 2.4 seconds longer in sentence edits that result in a correct sentence. During the course of the experiment, this difference may well mount up to a much larger overall difference between correct and wrong sentences.

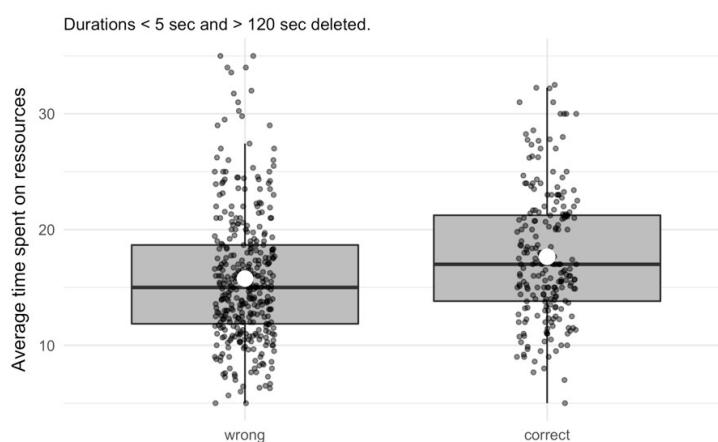


Figure 1: Boxplots for the time spent on resources for wrong and correct sentences. All times below 5 seconds have been excluded to exclude excessive switching between Excel and the browser. All times above 120 seconds have been excluded to exclude cases where participants interrupt correction of sentence A, work on sentence B (and C, D...) and then come back to the sentence A. The large white dots represent the respective mean values.

A closer look at these data indicated that the average time spent with a resource is related to frequent zapping between the resources, e. g. subject R-01 worked on sentence 1 for 3'5'', undertook 25 actions, which results in an average time of 7.65'' on a single resource (without correction time). See Table 2 for an overview of R-01's correction history.

Table 2: Timestamp, seconds per action, name of the resource and actions of participant R-01 for sentence 1 (not considering the time spent writing the correction of the sentence in the Excel file)

<i>Timestamp</i>	<i>Seconds/ Action</i>	<i>Resource</i>	<i>Action</i>
00:04:08	18	Google.it	opens Browser, opens Google, enters search string: "wenn significato"
00:04:26	2	Dicios.it	chooses the suggestion of Google, opens Dicios
00:04:28	2	Excel	opens Excel
00:04:30	10	Dicios.it	opens Dicios
00:04:40	4	Google.it	switches back to Google search results
00:04:44	3	Excel	opens Excel
00:04:47	7	Google.it	switches back to Google search results
00:04:54	3	Google.it	opens Google, looks for "leo"
00:04:57	13	LEO Dictionary	opens Leo, looks for "quando", changes the language to "it"
00:05:10	4	Excel	opens Excel
00:05:14	16	Google.it	opens new Tab, opens Google, enters search string: "costruzione con wann"
00:05:30	8	Deutsch Info	chooses the suggestion of Google, opens DeutschInfo, enters search string: "Frage kondizionale con "wenn""
00:05:38	8	LEO Dictionary	opens Leo
00:05:46	12	Excel	opens Excel
00:05:58		Excel	types correction
00:06:50	6	LEO Dictionary	opens Leo
00:06:56	6	Excel	opens Excel
00:07:02	2	LEO Dictionary	opens Leo
00:07:04	3	Deutsch Info	opens DeutschInfo, enters search string: "Frage kondizionale con "wenn""
00:07:07	17	Google.it	switches to the results of Google
00:07:24		Deutsches Institut Florenz.it	chooses the suggestion of Google, opens DeutschesInstitut, enters search string: "Congiunzioni: l'uso di "als" e "wenn" (quando)"
00:08:00	9	Excel	opens Excel
00:08:09	18	Deutsches Institut Florenz.it	opens DeutschesInstitut, enters search string: "Congiunzioni: l'uso di "als" e "wenn" (quando)"

Other subjects spent more time with the individual resources and achieved better results. Given this observation, we investigated the following factors: a) number of actions, b) number of search strings, c) average time spent on resources d) idea before the search¹ and e) accuracy of the sentences. The quantitative analyses suggest that the less time the students spent with a resource and the faster they switched between them, the more difficult it became for them to gain a clear idea

¹ This variable was coded by the annotators who transformed the screen recordings into the structured dataset (cf. Table 1) and is set to "true" if the participants expressed an initial idea of the problem before referring to Google or a specific resource.

of how to solve the problem. An exception to this rule is subject R-07, who undertook few actions but spent a relatively short time on the resources, namely an average of 14'6". The opposite of the general statement above would mean: The more time a participant spent on a resource and the less (s)he zapped between the resources, the better was the intuition before the search and the more sentences were correct. This statement cannot be confirmed consistently by the data; subjects with frequently correct sentences spent relatively more time on the resources (> 20') but did not necessarily have an intuition before the search for all their accurate sentences. For example, subjects R-08 and R-14 were able to correct several sentences even without any idea before the search. It can be deduced that, in addition to the average time spent with the resource and the intuition before the search, there must be other factors that predict sentence correctness. The concrete questions that arose were: What are the additional factors that play a role, and why are these factors decisive for a successful search and the correctness of the sentence? These questions cannot be answered by purely quantitative analyses. It was at exactly this point when verbal protocols came into play. This will be illustrated by examples from participant R-07². Example 1 shows the verbal protocol for the sentence *Obwohl sich der Junge beeilt hat, hat er die U-Bahn **verloren*** (Eng. "Although the boy hurried, he missed the subway"). R-07 was aware of the polysemy of the Italian verb *perdere* (Eng. to lose, to miss) (line 8-9), which means that she had already developed a certain language awareness; she knew that in combination with a vehicle like *U-Bahn* (Eng. "subway") the German verb *verlieren* (Eng. "to lose") was not correct and that a specific verb has to be used instead (line 11-13). She was aware of words belonging together (collocations) and she consequently searched for a specific word in the resources (cf. also example 3). This is how she avoided a word-by-word translation (*perdere – verlieren*), which, in our stimulus sentences, usually leads to interference errors. In addition, she knew various resources and opened an appropriate resource related to the search query. In order to find out the meaning of *verloren*, she opened PONS (lines 3-4); for the conjugation of *verpassen* she opened Reverso (lines 12-14). As we can see in example 1, she used linguistic strategies: in her search, she used a synonym like *Zug* (Eng. "train") for *U-Bahn*, which she thought was more prototypical, and synonyms for *verlieren* (line 10, 12).

(1)

- 1 So I read the first sentence [she reads the sentence] (*Obwohl sich der Junge beeilt hat, hat er die U-Bahn verloren*)
- 2 eh the section in bold is *verloren* so the verb *ehm* there are two sentences therefore one is the main clause and one
- 3 is the subordinate clause *ehm* so first I look for the verb *verloren* I prefer google chrome # *ehm* # generally I use the
- 4 online dictionary PONS because perhaps it shows also the context and the use of a word and also some examples
- 5 so therefore I search from German to Italian okay so *verloren* ## which means # okay I see it's the past participle
- 6 of the verb *verlieren* so lost the past participle '*perduto*' lost ok # so I lost the train even though the boy (*hat beeilt*)
- 7 I look it up because I'm not sure what does it mean so # *beeilt* (*affrettarsi*) okay (*sich beeilen*) so even though the
- 8 boy hurried he lost the train #perhaps the error would be that *verloren* is used in other contexts so I look for other
- 9 use contexts or a synonym of the verb *verlieren* so # I look up *verlieren* and it shows me (*perdere perdere la testa #*
- 10 *dispandersi*) okay In Italian I look for a synonym of the verb *perdere* in German so I set Italian German and look for
- 11 *perdere* okay so (*verlieren verlegen smarrire*) # *eh perdere il treno* for example it shows (*verpassen*) which means
- 12 that I can use *verpassen* instead of *verlieren* and *ehm* okay it shows me (*Zug*) so okay I use *verpassen* I will look
- 13 for the past participle for being sure so I open a website with the name REVERSO ## so conjugation German verbs
- 14 I search on google and it gives me REVERSO so I look for *verpassen* # so (*verpasst*) okay I write the sentence on
- 15 the right again *obwohl sich der Junge beeilt hat hat er die U-Bahn verpasst* [she finishes the sentence] okay I go
- 16 continue with the second sentence

2 The transcription symbols are based on the *Lessico di frequenza dell'italiano parlato* (De Mauro et al. 1993): # = short break; ## = longer break; <?> word is not comprehensible; sotto<categoria> = the word was interrupted, but the reconstruction was possible; [sie liest den Satz laut vor] = extra-linguistic comment, (*hat beeilt*) = student reads a word, expression or sentence found in the resource. The original verbal protocol is in Italian. For obvious reason we have provided the translation in English.

The participant showed good metalinguistic knowledge of German and a high level of language awareness throughout the reflection process. At the same time, the subject read the grammatical annotations in the resource carefully and took them into account when finding the solution, as can be seen in example (2). The interplay of all these factors resulting from the verbal protocols might explain the longer time spent on the resources:

(2)

1 [She reads the sentence] (*An unserem Forschungsinstitut ist Ihnen unsere Bibliothek 24 Stunden zur Verfügung*)
 2 *Verfügung* okay so *nel nostro istituto* I don't know the word so later I look for it # then *la biblioteca 24 ore*
 3 *Stunden 24 ore* of 24 I think ok so first I look up ah *Forschungsinstitut* because I don't know the meaning so ##
 4 *Forschungsinstitut* # okay there is no result ok so I look only for *Forschung* and there is *indagine* so I think in
 5 our research institute *ehm* # I think that I need to say perhaps *c'è* so perhaps *gibt es* [?] I don't know so I look up
 6 *Verfügung* ## and it shows *disposizione* so it means it is available and there is written (*jemandem zur Verfügung*
 7 *stehen*) so instead of *ist* I might use *stehen* because it means *essere a disposizione* so I put in # but it is also wrong
 8 (*zur Verfügung*) # *ehm jemandem essere a disposizione di qualcuno* so I write *an unserem Forschungsinstitut ehm*
 9 *steht ihnen* because there was *jemandem* which means dative yes *ihnen unsere Bibliothek 24 Stunden* and I don't
 10 change mmm # *zur Verfügung* [she completes the correction of the sentence] okay

R-07 connected language competence, attention, metalinguistic reflection, language awareness and dictionary use awareness to arrive at a good correction of the sentence. It is also very interesting that she often double-checked her correction proposals (example 3, lines 5-6), i.e. she changed the search direction and checked her hypothesis, although she was quite sure of the solution. This proficient use of strategies was also responsible for the high number of correct sentences of this participant.

(3)

1 [She reads the sentence] (*Wenn ich zur Schule ging habe ich viel Sport gemacht*) so when I went to school *ehm* I
 2 did a lot of sport I did a lot of sport in this case it's wrong *wenn e ehm* # because I think that eh you have to use
 3 *als* instead of *wenn* but I still try to find out if it gives me a few examples always some context of use okay so it is
 4 also used as conditional but it is not in this case in this case it is a temporal clause I believe yes because it is used in
 5 the past so every time when it is used when an action of the past is repeated often so I look up the sentence *quando*
 6 *andavo a scuola* and I have a look if it's used also from Italian # to German

Finally, the determination not to give up and to find a solution seems to be another decisive factor (cf. example 4). Other participants also had good metalinguistic knowledge, had opened a useful resource, had read the information given in the resource attentively and were close to the solution. However, they sometimes lacked the determination not to give up and to find a satisfactory solution. For example, after several unsuccessful searches (lines 1-4) and a certain insecurity (line 4), student R-14 gave up and went on to the next sentence (line 5):

(4)

1 Okay no # I can't find the solution so I usually solve it by looking for the sentence on Google and see if it is used
 2 but in this case it doesn't give me examples I can't find anything # I try to check again on LEO dictionary if there's
 3 something else mmm no okay I don't find don't think it is (*tauschen*) but there is (*umtauschen*) or (*gegen etwas*
 4 *tauschen*) mmm # I'm not sure so I don't know the use of these verbs I try to search on the internet on Google okay
 5 I can't correct it so I go ahead

As we have shown, the verbal protocols bring to light certain behavioral patterns and the reasoning of the participants. In our case, it could be shown that not only the average time with the resources

and an intuition before the search are responsible for a successful correction of the sentences, but also language competence, language awareness, and the (correct) use of strategies. Careful reading, paying attention to metalinguistic annotations in the resource and the determination to find the solution also play a fundamental role – factors that would not have come to light through a purely quantitative analysis. How much these factors interact or whether they might be arranged hierarchically or operationalized in a quantitative way remains to be explored.

3.2 Intuition and Focalization During Research

While analyzing the verbal protocols and the screen recordings, we observed that in some cases intuition and hypotheses formulated at the beginning of the correction task seem to influence participants' search behavior. The students initiate a search process to validate their hypothesis and miss relevant information found in the online resources. In what follows, we will give a detailed qualitative account of participants' search behavior when they show such focalization behavior. We will propose a schema based on the search patterns observed. As will be seen below, the verbal protocols and screen recordings play a crucial role in helping us to understand what may lead students to exhibit specific search behaviors.

We begin by tracing the search actions by a Portuguese participant while trying to correct the following stimulus sentence: *Er wohnt seit Jahren in Berlin und trotzdem verliert er sich immer noch* ("He has been living in Berlin for years and still gets lost"). We expected the participants (i) to identify that although the common polysemous verb *verlieren* means "to lose", the verb *sich verlieren* is not allowed in this context; (ii) to search for the correct reflexive verbs that fit the above context: *sich verlaufen*, *sich verfahren* or *sich verirren*: "*Er wohnt seit Jahren in Berlin und trotzdem verfährt/verläuft/verirrt er sich immer noch.*" From the verbal protocol it is clear that the student was not sure about the meaning of the verb *verlieren*. This led him/her to search the form "*verliert*" in Google Translate. The Google Translate result is the Portuguese verb "*perde*" (*perder* - "to lose"). The verbal protocol shows that the student correctly inferred that the equivalent Portuguese verb is the reflexive form *perde-se* (*perder-se*): "# *ahhh ele mora anos la e mesmo assim perde-se # / # ahhh he's been living there for years and still he gets lost*" (excerpt from the verbal protocol). All further search actions were directed to confirming that the correct verb in the given context is *verlieren* and that it is reflexive. Following this assumption, the student entered "*verliert sich*" in Google Translate and obtained the result "*perdido*", the Portuguese verb without the reflexive pronoun. According to the verbal protocol, the student expects Google Translate to output the Portuguese reflexive verb "*perder-se*". Since this is not the case, (s)he entered more context taken from the stimulus sentence in Google Translate "*verliert er sich immer noch*" hoping to obtain the reflexive form of the verb. Once again, the Google Translate result does not contain the reflexive form of the verb: "*ele ainda perde*". According to the verbal protocol, the student begins to question whether the reflexive pronoun is needed in the German stimulus sentence: "# I think that *sich* is not needed here #".

As the Portuguese equivalent *perde-se* does not appear in Google Translate and this does not correspond to the student's expectation, (s)he changed the language direction in Google Translate from Portuguese to German and added the reflexive pronoun to the Portuguese sentence: "*ele ainda se perde*". Google Translate outputs a German translation without the reflexive pronoun: "*er verliert noch*". As can be seen in the verbal protocol, this led the student to conclude that the German verb *verlieren* is not reflexive in the given context: "# exactly I think it doesn't need *sich* #".

The qualitative analysis of students' search behavior via the examination of the verbal protocols and the screen recordings has proven to be an important method for identifying search patterns common to a specific participant group and across participant groups. Regarding the German stimulus

sentence above, we observed that eight out of the 10 Portuguese participants assumed – based on intuition before initiating the search process or on hypotheses resulting from search actions – that the verb *verlieren* or *sich verlieren* is correct in this context. From this point onwards, these participants arrived at one of two hypotheses: (i) the problem lies in the reflexive pronoun – *verlieren* does not take a reflexive pronoun; (ii) the problem lies in the word order of the verb *verlieren* and the reflexive pronoun – “*verliert sich*” or “*sich verliert*”. With all further search actions, they tried to confirm or validate their respective hypothesis. Regardless of the resource used (Linguee, PONS Dictionary, Google Translate), search actions that take “*verlieren*” or “*sich verlieren*” as a starting point lead to unsuccessful results.

These qualitative observations provided us with enough evidence to formulate a focalization hypothesis search pattern that is illustrated in Figure 2.

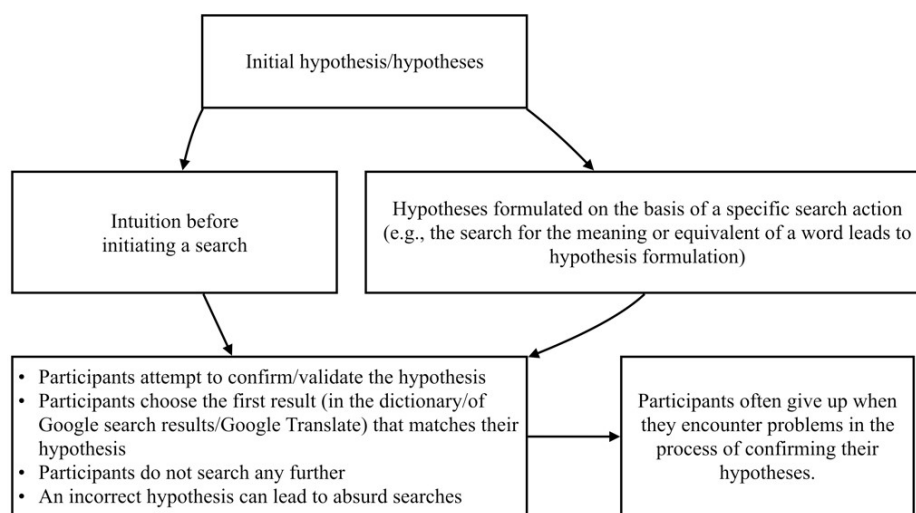


Figure 2: Schema of the focalization hypothesis search pattern

The students start off by formulating an initial hypothesis, based either on intuition before initiating a search process or on hypotheses formulated on the basis of a specific search action, such as the search for the meaning or translation equivalent of a word. The entire search process thereafter focuses on the attempt to confirm this hypothesis. The observational data seems to indicate that many students generally choose the first result they find in the resources that matches their hypothesis and do not search any further. As seen in the above example, an incorrect initial hypothesis more often than not leads to absurd search actions and results. We also observed that participants who encounter problems confirming their hypothesis usually give up on trying to correct the stimulus sentence.

In order to demonstrate the focalization hypothesis discussed above, we will provide a second example taken from the Portuguese observational data. In this example, students were asked to correct the German stimulus sentence *An unserem Forschungsinstitut ist Ihnen unsere Bibliothek 24 Stunden zur Verfügung* (“At our research center our library is at your disposal 24 hours a day”). Correcting the sentence involves identifying that the verb *ist* (Infinitive: *sein*) must be replaced by the verb *steht* (infinitive: *stehen*) in combination with *zur Verfügung* in this context. In other words, it is expected that students identify the function verb construction (German *Funktionsverbgefüge*) *zur Verfügung stehen*.

The student in question began by formulating the following hypothesis taken from the verbal protocol: “here the use of *ist* is not correct with *Verfügung* # I think it should be *gibt* instead of *ist* #”. Taking this hypothesis as a starting point, the student entered “*gibt zur Verfügung*” in the Google

search engine. The search engine outputs “zur Verfügung gibt” as the first link on the Google results page that in turn refers the user to the Linguee Dictionary. The student was so focused on validating her/his hypothesis formulated at the beginning of the search process that the simple fact that the expression “zur Verfügung gibt” appears on the results page suffices for the student to come to the conclusion: “exactly it is *gibt* that should be used because it is <...> correct #”. The search process ends at this point. The student does not select the link in the search result to access the information found in the Linguee Dictionary and does not search any further. Here, again, the exclusive focus on the initial hypothesis rules out the possibility for the student to arrive at other conclusions and following other search paths.

Although the above examples have been taken from the Portuguese participant group, the focalization hypothesis has also been observed in Italian and Spanish speaking participants. The qualitative analysis of students’ search behavior has allowed us to pick up on the focalization hypothesis. In future work we intend to complement the qualitative findings with quantitative methods in order to be able to compare the datasets in a systematic manner. The combination of qualitative and quantitative material will provide us with a more comprehensive insight into students’ search behavior.

4 Discussion and Summary

Given the two example studies above, we can draw a few tentative conclusions concerning the behavior of language learners when they have to resolve language problems in their L2. Generally speaking, the time that is spent consulting a resource pays off. The longer our participants stay on resources, the more likely they are to arrive at a correct final sentence. If more factors resulting from the verbal protocols are taken into consideration, the picture becomes more complicated but also clearer: thoroughness has to be accompanied by cross-checking preliminary conclusions (even when you are already quite sure about a solution), good meta-linguistic knowledge, and a strong determination to arrive at a good correction. The latter two factors are very hard to operationalize on a quantitative level. Hence, we presented evidence from the verbal protocols that allows us to infer these factors from the verbalizations of the participants. Another observation we presented in Section 3.2 is that many L2 learners start their research with a strong hypothesis in mind that guides their whole research process. This focalization, as we have called it, can be so strong that participants even ignore information that is readily available in the resource they are consulting. Alternatively, if the hypothesis guiding the search cannot be confirmed, some participants give up searching for a solution altogether.

Coming back to the main topic of this paper, we want to comment on some of the strengths of quantitative and qualitative approaches and connect those to our data. In the first example study, we started from the general quantitative observation that more time on resources leads to better results. This is a general pattern in our sample (or even larger groups of language learners) that can only be observed when the variables are clearly operationalized and analyzed by inferential statistics. Such pattern extraction and generalization on larger groups is a clear advantage of the quantitative approach. However, the qualitative approach allows us to complement the analysis with explanations for cases that do not fit the general pattern. By looking at individual search histories and the accompanying verbalizations, we can come to a more detailed understanding of the processes that help language learners to get to good solutions (or prevent them from getting to them). From there, we are able to generate further hypotheses that can be tested with our data in a quantitative way.

Finally, we want to stress that one of the main research questions of our study (What do language learners really do when they are solving language problems on their own?) can be answered in a way that is very encouraging for lexicography. In the vast majority of all sentence corrections (78.1%),

our participants used an online dictionary of some sort. Automatic translators (like Google Translate or the PONS Translator) were also widely used, but only in 21.9% of all sentence corrections. Given our sample, this suggests that language learners still rely heavily on lexicographic resources – at least on the web – even when they are allowed to use any resource they want.

References

- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. In *Psychological Bulletin*, 70(4), pp. 213–220.
- De Mauro, T., Mancini, F., Vedovelli, M., & Voghera, M. (1993). *Lessico di frequenza dell'italiano parlato*. Etas, Milano.
- Dziemianko, A. (2012). On the use(fulness) of paper and electronic dictionaries. In: S. Granger, M. Paquot (eds.) *Electronic Lexicography* (pp. 319-342). Oxford: Oxford University Press.
- Levy, M. & Steel, C. (2015). Language Learner Perspectives on the Functionality and Use of Electronic Language Dictionaries. In *ReCALL* 27(2), pp. 177-196.
- Lew, R. (2015a). Opportunities and limitations of user studies. In *OPAL – Online publizierte Arbeiten zur Linguistik*, 2/2015, pp. 6-16.
- Lew, R. (2015b). Research into the use of online dictionaries. In *International Journal of Lexicography*, 28(2), pp. 232-253.
- Müller-Spitzer, C. (ed.) (2014a). *Using Online Dictionaries*. Berlin/New York: de Gruyter. (Lexicographica: Series Maior 145).
- Müller-Spitzer, C. (2014b). Empirical data on contexts of dictionary use. In: C. Müller-Spitzer (ed.) *Using Online Dictionaries* (pp. 85-126).
- Müller-Spitzer, C., Domínguez Vázquez, M. J., Nied Curcio, M., Silva Dias, I., & Wolfer, S. (in prep.). The right hypotheses and careful reading make the difference: Results of an observation study with language learners on using language resources online.

Acknowledgements

We want to thank the Institute for the German Language (IDS) in Mannheim, Germany, for supporting our study with funds from the institution's core budget. We also want to thank all the assistants that helped transcribing the data and supporting us with their native language skills during data acquisition.