Corpus-Based Metalinguistic Discussions Designed as Writing Pre-Tasks

Eva SCHAEFFER-LACROIX
University Paris-Sorbonne (IUFM de Paris)
CoVariUs (Contextes, Variation, Usages), research group in linguistics of German and Scandinavian languages
Paris, France
elacroix@paris.iufm.fr

Abstract

A metalinguistic discussion is a form of oral or written verbal exchange about language features. Such discussions can be integrated in data-driven foreign language learning scenarios. Culioli, the founder of the theory of enunciative operations, considers corpora as knowledge resources which may facilitate the observation of variety in language use (Culioli, 2002: 131). Hence, it seems relevant to combine aspects of the conceptualisation approach (Chini, 2009), inspired by Culioli’s work, with concordancing activities, such as proposed by Boulton (2010) and others. This can be done within a task-based framework, taking into account pre-intermediate or intermediate learners who are not specialized in linguistics.

This paper analyses the effects of corpus-based metalinguistic discussions executed during a film script writing project called Filmlokal. The project took place during regular German lessons offered in May 2011 to two classes of 15-16 and 16-17 years old students in a French sixth form college. The levels ranged between what is called A1 (basic user) and B1 (independent user) in the Common European Framework of Reference for Languages. During the pre-task period, the learners were encouraged to think and to talk about the spatial meaning of German verb particles – e.g., "auf/stehen" [get up] – which were presented as relevant elements for the film script section called "stage directions". The particles could be identified in five film scripts which together formed a corpus I stored on my personal section of the corpus management system Sketch Engine (Kilgarriff, Rychly and Pomikalek, nd). The students had direct access to this automatically annotated online corpus. In addition, worksheets based on the multilingual Opus subtitle corpus allowed them to compare German and French concordance lines containing spatial elements. The students were invited to publish the results of their discussions on Edmodo, a virtual learning environment. I analysed audio recordings, screencasts, text productions and structured interviews. A certain progress in the domain of the number and the use of verb particles and in the domain of text organization can be attested for some of the students. However, the monolingual corpora was revealed to be challenging especially for basic users. A bilingual corpus and online discussion tools, such as Skype and TitanPad, could help to better take into account individual learning needs during concordancing sessions with a metalinguistic focus. Some of the results of the Weissensee project, a study with these characteristics, complete the Filmlokal data. This project is based on a home-made parallel corpus containing the film script of the series Weissensee. It was executed with a small group of future documentalists in spring 2012. The analysis of selected metalinguistic discussions of both of the projects supports optimistic conclusions about the benefits of the presented corpus-based learning scenario.
Introduction

In the eighties and nineties of the 20th century, researchers such as Bailly (1984) considered talking about language features as a relevant means of foreign language learning. At the same time, concordancing activities were developed by Johns (1988) and others. About twenty years later, these two forms of activities are coming back on the scene. It seems useful to combine both of them in order to share their potential for supporting foreign language learning: corpus data are likely to be exploited by learners in a way that leads them from experiencing frequency and other observable facts to restructuring of representations about language features. Two main questions guide the following presentation of the research data: what happens if the teacher decides to let concordance lines "speak"? Does the use of the concordancer, combined with metalinguistic discussions, lead to observable learning outcomes?

Metalinguistic discussions

Metalinguistic discussions are oral or written exchanges about language features. Their possible cognitive effects have been discussed by Chini (2009: 147) and others. They can be conducted by a teacher or a tutor in presence or distance learning settings. Questions arise about the way to organize the discussion, the proposed input, the metalinguistic vocabulary presented (or not) to the students, and the teacher's or tutor's guiding style. The content of the discussions may depend on needs expressed by the learners, or it could consist of language features considered as significant for the execution of learning tasks.

Focused pre-tasks

Tasks can be considered as a means to structure learning settings. Some existing studies explore the role of task types in foreign language learning: comprehension tasks, production tasks, or tasks with a language focus (focused tasks). The latter is of special interest for the studies presented in this article: "Focused tasks are tasks designed to provide opportunities for communicating using some specific linguistic feature (typically a grammatical structure)" (R. Ellis, 2009). In the same paper, R. Ellis evokes the possibility to include explicit grammar teaching in pre-tasks: "TBLT can include a pre-task and post-task phase, where opportunities arise for the explicit teaching of language". Together with Yuan, in a paper about oral production tasks, he describes observed effects of pre-task planning on grammatical complexity and on fluency and lexical variation in production (Yuan & R. Ellis, 2003).

(...) pre-task planning enhances grammatical complexity while on-line planning positively influences accuracy and grammatical complexity. The pre-task planners also produced more fluent and lexically varied language than the on-line planners (Yuan & R. Ellis, 2003).
Research questions

Yuan's and Ellis' statements above support the idea that the analysis of pre-task data could provide insight into the way learners gain from language observation during a writing project. The Filmlokal project attests a certain progress in the domain of the number and the use of verb particles and in the domain of text organization for some of the students. However, the use of monolingual corpora revealed as challenging especially for basic learners. Could a bilingual corpus and online discussion tools, such as Skype and TitanPad, help to better take into account individual learning needs during concordancing sessions with a metalinguistic focus? Can concordance lines, combined with metalinguistic discussions, function as a relevant knowledge resource and, more specifically, can the concordancer function as a writing aid? The following description and the analysis of the Filmlokal project data, completed by data from the Weissensee project, intend to provide some elements of response.

Theoretical and methodological background

Conceptualisation

The conceptualisation approach (Bailly, 1984; Chini, 2009) applies principles of Culioi’s theory of enunciative operations to language learning especially in high school contexts, but not restricted to those. The idea is to let learners observe linguistic data and to talk with them about their findings in order to question existing representations about language and to help to modify them, if necessary.

Spatial semantics

Chini (2009: 12) pleads for the taking into account the learning process and not only the learners’ productions. Hence, it seems crucial to offer activities which may help them to question their language representations. As space is conceptualized by different lexical patterns in German and in French, the field of spatial semantics could allow the observation of some aspects of construction processes. Krause (2004) considers particles and prepositions as a neglected chapter in German pedagogical grammars. Corpus exploitation could be a way to complete or even to review descriptions provided in existing workbooks. This sort of activity may help learners of German, and even their teachers, to get a better understanding of the role and meaning of spatial verb particles and prepositions.

Observed data

The collected data provide information about what the learners wrote and said during the project and about their query methods. Written traces are represented by the learners’ texts: two short film reports, a film script sample, findings of language observation sessions and posts published on Edmodo, Skype and TitanPad. Oral traces are supplied by recordings of verbal exchange (between the members of small learner groups, or between learners and the teacher) during the sessions about language observation and text revision. The learners'
actions on the computer were filmed with the screen capture tools Camtasia studio (video and audio), JMCFilmEcran (only video) and Screencast-o-matic (only video). The audio recordings and the films show, to a certain degree, how the learners did queries on the Sketch Engine server and how they published their results on Edmodo.

Projects Filmlokal and Weissensee

It has been about six years since I first integrated corpus exploration tools into writing projects. These projects were all carried out with French speaking learners of German as a foreign language, at high schools or at university. This article presents data from Filmlokal and Weissensee, two studies having a strong linguistic focus. Both of them are based on corpora containing German film scripts, and they invite the learners to write a small film script or a part of it, after a period devoted to the observation of spatial verb particles and prepositions. As a matter of fact, these linguistic features occur with significant frequency in the stage direction parts of film scripts.

The context of the projects

The Filmlokal project took place at the high school at the south of Paris where I worked as a German teacher. The project was carried out with a class of nine students, aged 15-16, and a class of twelve students, aged 16-17. All the students had been learning German for at least four years, and they were optimistically expected to achieve a level between what is called A2 (waystage) and B1 (threshold) in the Common European Framework of Reference for Languages (2001: 22). The Weissensee project took place at the Teacher training college IUFM de Paris. Three students in documentation studies were offered a 24 hours German course, based on the same principles as Filmlokal. Their language level ranged between B1 and B2.

The learning scenario

During eleven sessions in May and June 2011, five groups of four or five teenagers, built with students from the two classes, were given the opportunity to write a short film script sample. Before the tackling of the writing task, they were subjected to a pretest before observing together the meaning of selected particles in offline and online concordance lines. They discussed the findings with their peers or with me and they published the results on Edmodo, a virtual learning environment. A posttest and an interview with an external researcher in applied linguistics closed the project. The same procedure was applied in spring 2012 to three documentation students1, with the difference that they only had to write an alternative text for one of the film script scenes, and the metalinguistic discussions were organized in a more

---

1 Three seems a small number for a study. As a matter of fact, I consider the Weissensee research as one step on the further development of a corpus-based text revision method. It is planned to repeat this method later on with several other test groups.
planned and controlled way: they took place an hour per week, that is, more intensively and with more teacher intervention than during the *Filmlokal* project.

**The corpus management system**

The ideal corpus management system for genre-based projects like *Filmlokal* or *Weissensee* allows corpus owners to upload their corpora, to have the data automatically annotated and to do queries which allow the user to observe concordance lines, but also to enlarge progressively the context size, until attaining the source text. Last, but not least, it should correspond to the students' technical competences. After comparing and testing several tools, I chose the corpus management system *Sketch Engine* which has proven suitable even for basic learners. Its search options include grammar categories which may support metalinguistic discussions.

**The corpora**

Two online corpora were used during the *Filmlokal* project, the multi-lingual *Opus subtitle corpus* (Tiedemann, 2007) and the monolingual corpus *Drehbuch* I created in 2011, containing four film scripts of contemporary movies. I compiled the *Weissensee* corpus which is entirely based on a television series.

**Opus subtitle corpus**

In order to allow even a basic learner to observe the semantically challenging chapter of verb particles, I decided to provide the students with offline worksheets based on the German-French search interface of the *Opus subtitle corpus*. Due to the insufficient quality of the French text, I didn't recommend the learners consult the corpus directly.

**Corpus Drehbuch**

The corpus *Drehbuch* contains film scripts of four movies, published online by the editor *Verlag Zweitausendeins* (Frankfurt am Main). I used the corpus management *Sketch Engine* to compile a corpus containing 130,501 tokens and to have them automatically tagged with *TreeTagger*.

**Corpus Weissensee**

Like the corpus *Drehbuch*, the corpus *Weissensee* has been published on *Sketch Engine*, and it was automatically annotated. I compiled the complete script of a series called *Weissensee* (Hess, 2010) which contains 82,030 number of words. I translated the first part of the series which made it possible to offer an aligned German-French corpus on *Sketch Engine*, containing 13,640 German words and 15,832 French words.

---

2 I obtained permission to download all the scripts used for *Filmlokal* and *Weissensee* and to use it for research purposes.
Metalinguistic discussions, designed as corpus-based pre-tasks

Before planning and executing the writing task, the small groups were invited to execute pre-tasks, more precisely, to resolve together several language riddles, called "Sprachrätsel". The learners observed spatial verb particles – for example, "los" versus "weg" [away] – or spatial prepositions. To do this, they were provided with handouts containing parallel concordance lines pre-selected from the German-French Opus subtitle corpus or from the German-French aligned Weissensee corpus. The bilingual work was completed by the exploitation of monolingual occurrences directly accessed by the students in the corpus Drehbuch or Weissensee with the help of the Sketch Engine concordancer. The students of the Filmlokal project published their results on Edmodo, the virtual learning environment used for both projects. The Weissensee project included metalinguistic discussion sessions between the students and the teacher, based on oral exchange via Skype and/or written exchange via TitanPad, a collaborative online workspace. This time, the language riddles were based on enunciative operations of location applied by Gilbert (2005) to the three English prepositions in, on and at. The riddles were completed by the discussion of intriguing examples selected by the documentation students in the monolingual film script. During both of the projects, the film script writing task offered the opportunity to reuse what had been discovered about spatial verb particles and/or prepositions thanks to the pre-tasks.

Results in the domain of language observation

Verbal exchange with language experts seems to play an important role in language observation. This can be illustrated by selected metalinguistic discussions coming from both of the projects. They will be presented in relation with two themes I expect to have supported foreign language learning during the projects. The first theme about the identification of the nature of tokens is rather form-oriented, and the second one, more content-oriented, concerns the meaning of near-synonymous verb particles and prepositions.

The nature of tokens

Several discussions show how learners deal with questions about the nature of tokens, especially of spatial verb particles or prepositions. During the Filmlokal project, Elsa's and Philippe's query around the particle "los" [away] shows Elsa's difficulties in distinguishing it from the token "gehen" [go], one of the verbs which can be preceded by "los" (los/gehen). In the context menu, Elsa clicks on "particle", perhaps because "los" is a particle. She seems not to understand the function of this menu which informs about the words co-occurring with the query word. During the session, she asked her classmate if "*geh/loses" would exist, and

Figure 1. Corpus Weissensee on Sketch Engine.
Philippe informed her that she was wrong. This sample suggests that Elsa had not yet constructed enough knowledge about particle verbs. In fact, the particle is the last word in a main clause and it doesn't have verb endings.

The exchange between Hubert and me, based on the observation of the token "an" in Sketch Engine concordance lines, allowed him to understand that "an" can function both as a particle and as a preposition.

One of the discussions between the documentation students and me concerned prepositions co-occurring with the verb "sitzen" [to sit]. Solène proposed "als" [than] as a possible solution. She had identified it in the example "besser als im Warmen zu sitzen" [better than sitting at a warm place]. Marjorie explained to her that "als" is not a preposition, and I gave an example and its translation in order to show that it is a conjunction. During one of our TitanPad discussions, Marjorie described her difficulties to deal with the double nature of "auf" which can be a verb particle or a preposition.

Near-synonymous verb particles or prepositions

The double meaning of "vorbei" [passed]

The guided OPUS concordance lines observation session with Elise, Jeanne, Martine and Valentin led to following result: "vorbei" [passed] has as well a spatial as a temporal meaning, and its spatial meaning is defined as "passing along something".

The difference between the meaning of "los", "weg" and "ab" [passed, away]

Both projects contained a language riddle which required to think about the difference between the verb particles "los" and "weg", completed by "ab" for the Weissensee project. The learners of both cohorts were able to detect that "weg" makes the departing element invisible and that "los" indicates that an element which was originally sort of fixed is liberated. Elsa and Philippe started with the hypothesis in mind that "los" would mean "quickly". To begin with, I found this observation non relevant; I had rather expected them to identify the meanings "leave" and/or "lose contact". Later on, the same riddle solved by Cornélie, Marjorie and Solène, the documentation students, helped me to understand that Elsa and Philippe had detected one of the essential semantic features of "los". As a matter of fact, Cornélie, too, perceived quickness in examples containing this particle. This is shown in the following Skype chat sample, referring to a fill-in-the-blank activity.

At the 11th [line], I put los. (...) Because the car starts quickly⁴.

In addition, Cornélie noticed a certain "brutality" in the bilingual concordance lines containing "los", a remark which gave me the idea that a rocket-like vectorial energy would characterise verbs starting with this element. This seems to distinguish it clearly from its near-synonymous particles "weg" and "ab".

⁴ I translated all the cited examples from French or German to English.
The verb particle "ab" was defined by students of one of the Filmlokal groups and by the Weissensee group in a way which surprised me first. Flora, Sylvain, Marc and Pia wrote the following observations on the Edmodo assignment space:

"Ab" has got a notion of removal (with "break" or "take away"); or a notion of stopping (with "defuse" or "cancel"). The particle is able to adapt itself to humans or to objects.

In the following Skype chat lines, Cornélia insisted on the idea that "ab" would stop something on-going (an action, or an event?).

[4/2/2012 7:05:07 PM] Cornélia: I would say "interrupt"
[4/2/2012 7:05:47 PM] Cornélia: A change in continuity perhaps?
[4/2/2012 7:06:07 PM] Cornélia: for "jump", "start"
[4/2/2012 7:07:12 PM] Cornélia: That is because she jumps off the swing after swinging⁵

Initially, both of the observations, "notion of stopping" and "change in continuity", seemed strange and not relevant to me, but after due reflection, I understood that the students were probably right: "ab" can be described as a two-way marker which indicates the end or the beginning of actions or events. The verbs "jump" and "start", cited by Cornélia, reinforce this idea.

Conclusions

This paper dealt with the following questions: can concordance lines be used as a relevant knowledge resource and thus be exploited for writing tasks? Does the expected benefit of the concordancer depend on a certain form of corpus-based metalinguistic discussions between learners and/or between learners and a language expert? It seems, indeed, possible to rely on the relevance of students' concordance line observations about the meaning of elements which are sometimes tricky to describe even for linguists, let alone for foreign language teachers. The pre- and posttest and the text productions of both of the projects show encouraging results: the number of relevant verb particles and prepositions tends to increase in the productions of a good part of the participants, and a better text organization can be attested for some of them. The next step to do is to exploit the data with respect to the outcomes of corpus-based metalinguistic discussions in the domain of text revision and obtained text quality.

⁵ Cornélia is referring to following input: "Auf einem Spielplatz am Ufer schaukelt ein 6jähriges Mädchen (…). LISA: Jetzt springe ich ab." [On a play ground near the water a six year old girl is swinging. LISA: Now I am going to jump.]
References


Tools


Tiedemann, J., (nd), *OPUS Open subtitle corpus*. http://opus.lingfil.uu.se/bin/opuscqp.pl