

Tool mediation in Focus on Form activities: case studies in a grammar-exploring environment

PETTER KARLSTRÖM, TERESA CERRATTO-PARGMAN

Department of Computer and Systems Sciences, Stockholm University/Royal Institute of Technology, Forum 100, SE-164 40 Kista, Sweden
(email: petter@dsv.su.se, tessy@dsv.su.se)

HENRIK LINDSTRÖM, OLA KNUTSSON

School of Computer Science and Communication, Royal Institute of Technology SE-100 44 Stockholm, Sweden
(email: henrik.lindstroem@gmail.com, knutsson@csc.kth.se)

Abstract

We present two case studies of two different pedagogical tasks in a Computer Assisted Language Learning environment called Grim. The main design principle in Grim is to support 'Focus on Form' in second language pedagogy. Grim contains several language technology-based features for exploring linguistic forms (static, rule-based and statistical), intended to be used while writing. Our question is, in what ways does Grim support Focus on Form in actual classroom use. We have explored this question within sociocultural theory, emphasizing tool mediation and how tools shape the learner's activity. The first case concerns a text-reconstruction exercise in which students worked in a pair within the Grim environment. The second case was conducted with another group of students, who engaged in collaborative revision of texts, written in advance by one of the students, in student pairs. In both studies, students were instructed and encouraged to use the different features of Grim. Data was collected by recording dialogue during the sessions with Grim. Our results show how learners put the features of Grim into use in their writing tasks. In some instances, the program was used creatively, in combination with external tools such as the users' own dictionaries, knowledge of other languages, or teachers. In other instances, we note that Grim was used for error correction, rather than as a language resource. The learners' activities are thus transformed by their use of the program, from the tasks of revision and text-reconstruction into error correction. The application shapes the activity, in conjunction with the pedagogical tasks. We argue for studying the activities of students with CALL tools, in order to find out in detail how tasks and technology concur in use and what view on language and pedagogy they mediate.

Keywords: Sociocultural theory, mediation, Focus on Form, collaborative learning, language technology, natural language processing

1 Introduction

It is perhaps a truism to state that tools are fundamental for human existence. Yet, for this very reason, it is quite important to find out in detail how they affect us. In its deepest

sense, the question of how we use tools concerns who we are.

The digital computer is a recent addition to the range of tools that humans use. Nevertheless, it has already had tremendous impact on our lives, in what is often called the “information age”. Another tool for information, language, has been an integral part of humankind for as long as we have been humans, if not longer. Still, as the histories of subjects such as philosophy and linguistics show, branding the inner workings of language an evasive problem is perhaps an understatement. Questions about what human language fundamentally is, and its relation to thought and reality, have resulted in countless debates over thousands of years.

Computer Assisted Language Learning (CALL) is a subject that simultaneously has to deal with age-old philosophical questions and rapidly changing technological developments. As CALL is interdisciplinary, we need to draw upon theory from several different areas, some of them themselves interdisciplinary: Linguistics, Pedagogy, Philosophy, Psychology, Second Language Learning/Acquisition (SLL/A), Language Technology (LT), Computer Science, and Human-Computer Interaction. Perhaps, then, it is no wonder that the search for paradigms in CALL has been subject to debate (e.g. Chapelle, 1997,1999; Salaberry, 1999). In one attempt at grasping this diverse subject, we have chosen to frame our research in sociocultural theory (SCT), emphasizing tools, how those tools mediate reality, and what activity people participate in when using specific tools.

Within SCT, activities are not the same as tasks, and cannot be designed in the sense that teachers or computer applications can provide them. Tasks, on the other hand, may be designed, in the hope that they will stimulate activity that is beneficial for learning a language. Activities are a central concept in SCT. To dare a simple definition, they are purposeful, transformational actions. They essentially form reality, the arena in which human conduct takes place (Lantolf & Thorne, 2006: 215). Learning a language is in that sense about people participating in activities that transform them into playing new parts in that arena.

Studies within the paradigm of sociocultural theory focus on authentic activities, and these are approached with theoretical grounding in the works of Vygotsky, Luria, Leont’ev, Bakhtin, and their followers. Research is based on theory-guided observation and interpretation of people involved in different kinds of activities (Wertsch, 1991; Lantolf, 2000).

Perhaps the most distinguishing feature in SCT lies in the strong emphasis on the fact that the relationship between mind and nature is bidirectional. Biology forms the foundation for thinking, but higher forms of thinking are culturally mediated. Nature, culture, and history affect our minds and vice-versa. Wertsch (1995) defines the goal of research in Vygotskian sociocultural theory to be to understand human mental functioning and its connection to cultural, historical, and institutional settings. Apart from the aforementioned *activities*, four cornerstones of this theory are the *genetic method*, *mediation*, *internalization*, and the *zone of proximal development*. (Lantolf, 2006: 18) Our perspective is on tools, material as well as psychological, although the focus on computer tools is primary in our work. In the two studies presented here, we have therefore taken tool *mediation* as our key concept, but the other cornerstones also influence the activities we observe and how we observe them.

Vygotsky’s view on *mediation*, then, relies on the observation that tools/artifacts,

concepts and the activity itself always mediate human activity and reality. The *genetic method* applied to tools entails the historical study of how the tools were developed, how they are transformed by use, and the transformative power of tools on human consciousness (Haas, 1996, 1999). Pedagogically speaking, the *zone of proximal development* conceptualizes learners' emerging capacities together with teachers and/or peers in the use of language and other tools. The fourth cornerstone is *internalization*, in which interpersonal and person-environment interaction forms and transforms humans' internal mental functions.

Sociocultural theory is gaining in popularity in Second Language Learning because it can be used to capture what individual learners do and who they are, rather than attempting to find a universal method for what they should do (Lantolf, 2001). It is an approach where questions regarding individuals and their use of internal and external tools are central. CALL is fundamentally about two tools, language and computers, and we believe that our area will benefit greatly from taking a similar approach.

In work closely related to ours, Gánem Gutiérrez (2006) suggests that one use of SCT in CALL is to investigate what impact the machine has in collaborative activity. Bearing that in mind, we are interested in how a CALL tool is put to use, what kind of activity it will encourage, and what view on language and education it mediates. Our question, then, is in what manners students put a particular tool to use in a particular context. In other words, how do the students interact with the tool, and what is the role of the tool in their dialogue about their writing? We investigate these questions in an explorative and descriptive rather than a controlled and prescriptive way. By moving our studies into classroom-like tasks rather than moving students into controlled experiments, we hope to capture real use of tools as accurately as possible.

2 Pedagogy and technology

The understanding that grammatical forms are complex and have both cognitive and communicative dimensions is important in sociocultural theories of SLL (DiCamilla & Lantolf, 1994). Our interest in designing language tools for writing relies on the central place that writing occupies in the development of language and thinking processes (Vygotsky, 1962, 1978). As Olson (1995) suggests, writing does not only allow us to do new things with language, such as mass distributing and storing meaning. It also turns language into an object for reflection and analysis.

Thus, writing provides a model, useful for gaining a common understanding of language structure and for explaining some of its mechanisms. This model must not be confused with a complete explanation of language, or worse, a collection of prescriptive rules to be obeyed at all times. Instead, in Swain's (2000) terms, writing could be regarded as a base for hypothesis testing. It invites the learner to reflect on language, and thereby become aware of grammatical forms and how they influence expression of meaning.

2.1 Focus on Form

Within more mainstream SLL theories there is a pedagogic concept called "Focus on

Form” (Long, 1991; Doughty & Williams, 1998). The basic idea behind Focus on Form (FonF) is to draw attention to linguistic code features when the learner is already involved in communicative tasks. This distinction between other forms-based pedagogy and Focus on Form is an important one. Focus on Form advocates communicative tasks that incorporate reflection on forms, rather than exercises that are concerned with forms only.

Focus on Form emerged in the early 1990s as an attempt to balance the sometimes infected debate between behaviorist and innatist views on language learning. In the behaviorist view, language learning was considered to take place primarily by practising linguistic forms by means of drills. What Long & Robinson (1998:18) call an “equally single-minded focus on meaning” builds to a large extent on Chomsky’s (1965) suggestion that grammar is innate. Because there is no need for children to learn grammar when acquiring their first language, there should be no need for grammar when acquiring a second one. Instead, students are expected to learn language by receiving “comprehensible input” (Krashen, 1985).

These are of course rather polarized views that perhaps belong more in academic discussion than in real classrooms. Focus on Form balances these opposing views by reintroducing grammar in communicatively oriented research and curricula. Thus, it might be said to be less polarized, and more practical with respect to what teachers do in class. Still, real classrooms, including the ones where our informants participated, are not purely FonF either. Instead, they contain different modes of teaching in different phases of the students’ learning.

Here, our use of FonF is somewhat relaxed. We do not intend to use it in its strictest sense of learning language while being taught another subject, or as leverage to reject grammar/vocabulary drills in CALL. FonF is not the one and only overarching pedagogy in our informants’ classes, but it is a feature of individual tasks. We merely

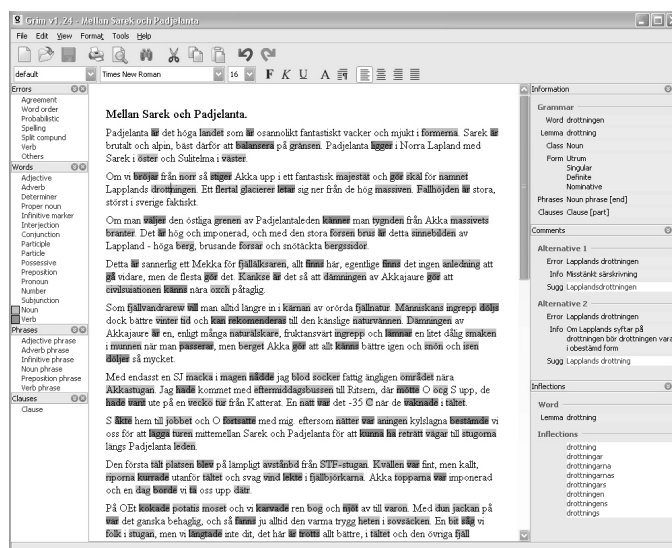


Fig. 1. Grim, here shown highlighting verbs and nouns.

suggest that tasks involving FonF may be useful in CALL, and wish to investigate their usefulness in more detail.

2.2 The tool – Grim

The computer application we have studied, Grim, is a multi-purpose grammar-exploring tool. The main ideas are to give the learners feedback on form and to provide them with tools for exploration of their interlanguage and the target language, Swedish (Knutsson, Cerratto-Pargman, Severinson-Eklundh & Westlund, 2006).

In Figure 1, Grim is shown highlighting nouns and verbs in a document. As can be seen in the left hand pane, other word classes may also be highlighted for similar purposes. This feature is meant to be used to reflect on word order, verb inflections, tense changes, etc., for example when revising a text. Other selections in the left hand pane are used to highlight phrases, clauses, and errors. It is possible for students to highlight different types of errors, in order to consider one particular error category at a time. The checking engines are rule-based as well as probabilistic.

In the right-hand pane, different kinds of linguistic information are displayed. Here, the grammar and spelling checkers give suggestions for handling errors. In addition, possible inflections are shown and various grammatical information is provided.

Finally, in the menu, three further features are available: a concordancer, a dictionary containing eight bilingual lexica, and a text summarizer. The lexica are enhanced by Language Technology (LT) to score more hits. The concordancer and dictionary provide further linguistic information, while the summarizer is more useful when working with shortened versions of others' texts.

Grim is to a large extent built on LT. In a sense, it is a reaction to LT having fallen out of fashion in language learning communities during the last two decades (Borin, 2002; Nerbonne, 2003). Reasons for the current lack of LT is that teachers and pedagogues have judged earlier applications to be too expensive, faulty, and normative in their linguistic output (Vernon, 2000: 333), and that some learners and teachers have had unfulfilled expectations in the possibility of human-like properties in technology (Nerbonne, 2003). Furthermore, LT in the form of corrective checkers was regarded as belonging to unfashionable pedagogies at a time when communication took precedence over grammar (Chapelle, 2001: 8; Vernon, 2000:333). As we have briefly discussed above, modern theory advocating Focus on Form in SLL is more balanced, and includes both grammar and communication. In addition, LT has progressed to provide better engines, and influences from current human-computer interaction may yield applications that provide feedback in ways quite different from earlier systems.

While the area is perhaps somewhat underdeveloped, LT in CALL has seen some progress, and the current number of workshops and special interest groups (see e.g. EUROCALL NLP SIG and CALICO ICALI) suggest that it is again an emerging strand in CALL. Gamper & Knapp (2002) review Intelligent CALL prototypes that are on the leading edge technologically. Many of these contain components from LT. In the review, it is lamented that most of the systems remain in a prototypical stage, and that they are rarely used in real language learning.

We believe that the LT tools on display in Grim are prime candidate technologies for

CALL in Focus on Form curricula. Still, we need to carefully consider how the tools may be used in learning environments. It is, for example, not self-evident that spelling- and grammar checkers are necessarily the best implementation of LT in teaching. The authority of grammar checkers can and should be questioned. They may be inaccurate, while also suggesting an absolute view on language (Vernon, 2000). Therefore, we are interested in the possibilities for Focus on Form in Grim. Not all of these opportunities are corrective regarding forms, and some of them complement each other to yield hypothetically less normative results.

Instead of being solely corrective, the prototype was intended to provide users with features to stimulate and encourage reflective thinking. The intention was to design tools and language feedback for learners to explore, analyze and reflect on the written language they use. Of course, to design language tools for reflective thinking is troublesome because it implies creating features and language content able to challenge the inertia that inclines learners to accept suggestions at their face value. Learners must be willing to endure a condition of mental unrest and disturbance, to be uncertain and inquisitive. This is difficult for any learner, or, in Dewey's (1991) words: "Reflective thinking, in short, means judgment suspended during further inquiry; and suspense is likely to be somewhat painful. [...] To maintain the state of doubt and to carry on systematic and protracted inquiry - these are the essentials of thinking" (*op. cit.*: 13)

In our case studies, we will see that Grim does encourage reflective thinking in the shape of Focus on Form, but also that the tools in Grim steer the task into an activity of sentence-level error correction.

3 Case studies

We present two case studies where students participated in tasks which involved working collaboratively with texts. The first study concerned text reconstruction, and the second one text revision. The exercises were specifically made for FonF with support from Grim.

3.1 Methodological approach

We were interested in how our participants discussed language while conducting their tasks, and in particular how Grim shaped those discussions. To this end, we recorded their dialogue during the sessions on video (study 1) and audio (study 2) tapes. Then, we transcribed the raw data, and identified "language related episodes" (Swain, 2000; Lindberg, 2001). Language related episodes are those where students discuss their interlanguage. Following Swain, we do not limit the kinds of episodes searched for to those that contain metalanguage (i.e. explicit discussion about forms and grammar). The broader term metatalk concerns any kind of discussion about language.

There are two kinds of episodes: Form-based and Lexical. The first concerns grammar, phonetics, orthography, and interpunctuation, while the second concerns lexical words and phrases. Since we are interested in the interactions humans perform with a computer tool, we have added the category Tool Interaction to these two. Tool Interaction episodes consist of talk and interaction about and with a computer tool.

Some episodes belong to more than one category, e.g. when feedback from the tool encourages the students to discuss a grammatical change in their text.

The episodes we identified were used to interpret the activity that the students participated in, in particular how the tool was used. Therefore, the episodes that interested us the most were those containing Tool Interaction. By isolating the episodes where language was discussed while output from the tool was used, we aimed to pinpoint how the tool encouraged students to Focus on Form. We also made some overarching observations regarding the students' writing process and teacher comments.

3.2 Study 1 – Reconstruction

Two adult learners, whom we will call "Ali" and "Natasha", participated in this study. They were medical doctors who had migrated to Sweden and wished to pursue their professional careers here. In order to do that, they had to master the Swedish language. They were highly motivated and interested in learning Swedish. They spoke different mother tongues (Russian and Persian), and were therefore obliged to speak Swedish with each other. The course they attended was on Swedish at an advanced level, tailored for medical practitioners.

During our study, the learners participated in three text-reconstruction sessions, of which the first two were pilot sessions in order for them to learn the application and for us to adjust our recording techniques. The text-reconstruction tasks were constructed by us together with the teacher of the class. It was a kind of task they were familiar with, and suitable for Focus on Form with Grim. A story concerning a fictional patient was read to the learners, and they were asked to collaboratively reconstruct that story in their own words. Prior to the exercises, they were given instructions on the features of Grim, and they were encouraged to use as many of the features as they saw fit and to ask questions about Grim as they arose. We were present to answer questions about the tool during all three sessions.

Instructions concerning the tasks were given orally before each of the sessions. The students were asked to listen, take brief notes, and then retell the story in their own words. Before the sessions started, they were informed of limitations in linguistic engines, and therefore agreed to leave error markings unattended when they thought Grim gave an erroneous report. They were also asked to make final adjustments, disregarding Grim's reports, after they had finished. Before the second session, they asked us a question concerning homonyms and the difficulty for a computer program to analyze them, thus displaying rather deep insight into the limitations of computer programs, as well as knowledge of linguistic code features that might appear in the program.

Data regarding tool interaction during the first two sessions concern learning Grim's features and limitations, while the third session was more focused on discussion when writing and revising their text. The majority of tool interaction episodes from the first two sessions therefore consist of questions to us, while they worked more on their own during the third session. For that reason, the last session was the most interesting, because that was when they had formed their own use of Grim, appropriated the tool to meet their needs and expectations.

We will illustrate our findings with excerpts from our transcriptions, chosen from the collection of language related episodes concerning tool interaction. The excerpts are translated into English from Swedish, and somewhat abbreviated for space and clarity.

3.3 Results from Study 1

The text-reconstruction exercise filled its purpose insofar as students participated in metatalk concerning their text. During the last session, the proportion of episodes concerning form was 60% while lexicon and interaction episodes each had a ratio of 20% to the total number of episodes (The sum of 100% is coincidental. As mentioned above, some episodes belong to more than one category). This means that the majority of episodes regarding form took place without consulting Grim. Nevertheless, Grim had a large impact on these students' writing because getting rid of error reports became a major goal. During the exercises, Natasha was the one typing, while Ali read from his notes. They occasionally consulted Natasha's notes as well.

In all three sessions, the students conducted their work in three distinguishable phases after having been told the patient's story. First, they wrote the story from the beginning to the end, while reviewing their notes. Then, they located all constructs that Grim had marked as erroneous and corrected them. During the first two sessions, we asked them to make any final revisions, to attend to issues in the text that had not been captured by Grim's grammar checker, including stylistic issues and content. For that task, we suggested that they make use of tools other than the grammar checker. During the third session, they made final revisions to make sure that verb tense was correct. They did so without instruction from us, but also without making use of the tools in Grim. It is notable that they did not reorganize the text, make any overarching changes to tense for stylistic effect, or otherwise make stylistic changes during any of the revision phases. Instead, all changes tended to take place on a sentence-by-sentence basis.

This mode of work stands in stark contrast to how the teacher of that class usually conducts his work. He informed us that when teaching at the advanced level of these students, he usually starts out by commenting on overall stylistic issues in the text, and only after that comments on individual sentences. One reason why the students worked in the opposite direction may be that the immediately available grammar checker in Grim comments only on phrase level. Other tools such as word class highlighting could be useful for higher-level revision, but these were not used.

During the first session, they asked us about the usefulness of syntax highlighting. We explained how it worked, and gave them examples of when and how it might be useful. The example in question concerned highlighting pronouns in order to visualize to whom they referred in the text. They agreed with our explanations, and told us they found the feature useful, but they did not use it.

In many senses the application was treated in the way intended in the task, namely to discuss and reflect on linguistic forms while writing a communicative text. There are episodes where an error is correctly marked as an error, and the students are helped by Grim to detect it:

Excerpt 1. Session 3: Lexical, Tool Interaction

Natasha: eh anna can manage herself "fast" "fäst" ((Fast = in spite of (conjunction). Fäst = fasten (verb)))

Ali: "fast" not "fäst" ((points at screen)) "fast"

Natasha: "fast" her joints ache.

We mention this to establish that Grim's grammar checker does work to help students detect errors and correct them. However, dealing with faulty output from the grammar checker, and using other features in Grim, is more problematic.

By the third session, the students had ceased to use any feature other than the spelling and grammar checker, despite them knowing of other features and being encouraged to use them. In the very beginning of the session, they agreed between themselves to write the text first, and then use Grim while revising. In the following excerpt, Natasha is about to correct the first sentence they have written, which has been marked as incorrect. Ali stops her and Natasha asks what to do:

Excerpt 2. Session 3: Tool Interaction

Natasha: how do we?

Ali: but we can write the entire text / and then [correct] it.

Natasha: [then?] and then try and correct

They followed this agreed mode of conduct throughout the exercise, ignoring error alarms that occurred as they wrote. Then, they revised the text by working through sentences to get rid of any error markings. After that, they scanned through the entire text again. The last scanning was triggered by a discussion concerning verb tense. The problem they spotted was a discrepancy between past and present tense, the patient *was* taken to hospital, but she *has* problems with her joints:

Excerpt 3. Session 3: Form

Natasha: but we start in present tense and then move to [...] preterit [...]

Ali: read through it all ((laughs))

Ali: be, she is alive, that is why we must use present tense there.

They read their entire text aloud, stopping occasionally to discuss verb tense. This could have been an opportunity to highlight all verbs in the text to see where they switch between tenses. That opportunity was not taken. Instead, they made one correction not caught by Grim, but missed several others. They did not introduce any more error markings, even though the exercise permitted them to do so if they thought their version was better than the computer's.

We interpreted these actions to mean that they were working towards getting a result that was clear of error markings, rather than a text that told the story in their manner. We asked them about this after the first session, and their reply was, perhaps unsurprisingly, that the text "looked better" without the error markings.

The most striking example of trying to "please" the system is an episode where they (correctly) judged the system to give them faulty feedback. The (correct) statement "then she received antidepressants" was marked as incorrect. They worked hard to get rid of the error marking:

Excerpt 4. Session 3: Form, Tool Interaction

Natasha: erroneous [construct]
 Ali: [construct]
 Natasha: then she received [antidepressants]
 Ali: but eh we can eh change then to since then maybe ((Natasha types, the new sentence is still tagged as erroneous))
 Ali: mm // she took
 Natasha: received
 Ali: it is correct received / here took ((points)) hmm
 Natasha: good idea took
 Ali: we can change verb

First, they tried changing the preposition “then” to “since then”, a change in meaning as well as form. That change did not rid the screen of the error alarm, so they switched the verb to one acceptable by the system, but grammatically the sentence remained the same. The verb “receive” would have been a stylistically better choice because it implies medication prescribed by a doctor. They probably knew this, at least they knew that “received” as written was correct. Yet, they switched it to “take” to get rid of the error marking. The error marking has made them change the meaning of the sentence twice, but to please the system rather than to improve on the text.

We note that they indeed discussed linguistic form with the help of Grim, but also that they worked towards getting a positive response from the system rather than writing a text that is acceptable to them. They acted reactively to Grim’s error reports to make efforts in sentence-level correction.

3.4 Study 2 – Revision

Four students participated in the second case study. They were exchange students at a technical university in Sweden, and took part in courses in Swedish as a foreign language. Two of them, “Ulrike” and “Stephanie”, had German as their mother tongue, and studied Swedish at intermediate level. The other two, “Maria” and “Jean” had French and Spanish as mother tongues, and studied Swedish at advanced level.

The students’ task was to revise texts together, in pairs, with support from the features of Grim. They were informed about limitations in the linguistic competence of computers, and asked to improve on the text in a general sense. Each student brought one text he or she had written in a previous assignment. They participated in two sessions, switching texts between sessions. In addition to Grim, they were encouraged to use other tools that they would normally use when writing or revising texts, such as grammar books and lexica. They were instructed to speak in Swedish and mostly did so. The two Germans spoke German with each other from time to time, especially when discussing problematic forms.

3.5 Results from Study 2

Our data consisted of audio uptakes of the dialogues between the users during the sessions. The dialogues were transcribed and analyzed with focus on the different episode types: lexical, form, and interaction.

The distribution of episodes differed from the first study in that these four students participated more in tool interaction. Sixty per cent of the episodes concerned interaction about and with Grim. The proportions of episodes containing discussion about lexicon and form were similar to the first study. Twenty per cent of the episodes were lexical ones and 70% concerned form.

We noted that the German students participated in a higher ratio of form-based episodes, 90%, compared to 60% of the second pair. It is also notable that both pairs, in particular the Germans, participated less in tool interaction during the second session. This is to be expected since they were more familiar with the program during the second session: there was no longer a need to discuss basic interaction.

As in the first study, the alarms played a salient role in the interaction between the users and the program, and shaped the task into one of error correction. For example, the students did not revise the text as a whole, but only noticed the forms that were marked as false by Grim. This was especially apparent during the first session with each pair.

The first excerpts show how the users relied on the alarms as indicators of what was correct and what was not. In excerpt 1 and 2, we see how the students viewed the alarms as a measure of the quality of their texts. In 1, an excerpt from Ulrike and Stephanie's first session, Stephanie exclaimed "tadaa" when they were done with the last alarm in the text. She concluded, in German, that they were ready: "fertig". In the second excerpt, 2, from the first session with Maria and Jean, Jean evaluated his text according to how many alarms the program had produced, and how many of these he considered erroneous or mere slips or mistakes. They also commented on the length of time it took for them to look over the marked forms in the text: "that was quick".

Excerpt 1. Ulrike and Stephanie, Session 1: Tool Interaction

Stephanie: tadaa ok / fertig ((German))

Excerpt 2. Maria and Jean, Session 1: Tool Interaction

Jean: aa *okay* there are three more ((alarms))

[...]

Jean: *okay that was it*

Maria: *that was quick*

[...]

Jean: okay so mm / but it's good because I had three typos (xx) maybe two errors yes (xx) three more errors hu / even if I had read through it once more (xx) mhm

The learners used the alarms as a way of judging the text. There were only two modes: a marked form was erroneous, whereas an unmarked was correct. In excerpt 3, Ulrike and Stephanie struggle with a formulation that contains a number of errors. They used trial and error and when the alarm disappeared, they considered the new form as being correct without much hesitation. The new form however, was also incorrect.

Excerpt 3. Ulrike and Stephanie, Session 1: Tool Interaction, Form

Handled form: "[...] lära vi oss denna funnit regelmässigheterna [...]" (we to learn this found regularities)

Ulrike: (xx) spell error / no (mhm) but also possible mis- ach so //

aha and now

Stephanie: denna funnit (yes) / no fannen (funnen) LAUGHTER // funnen
regelmässigheterna finished analyzing (ha) ha cool (aha) /

The three excerpts above are from the first sessions with the two pairs. When investigating the second session we saw that the students had developed a more incredulous attitude towards the computer and its feedback. Between the first and second session, they received input from the teacher, input that they could compare with the corrections made by Grim. By doing this, they realized that the program had made some false alarms, and missed a number of important faults. In excerpt 4, Jean makes a remark about the program not detecting some errors, and that they should revise the text once more to look for forms that they themselves consider being faulty or worth discussing for some other reason.

Excerpt 4. Maria and Jean, Session 2: Tool Interaction

Jean: okay (mm) we have to go through the text once more (okay) in
order to find that / which the *program* doesn't / doesn't see

The activity in the second session was still to a large extent shaped by the alarms that were given by the computer. The overall schema for their revision was to deal with the marked forms, in linear order. Only after that did they revise the texts more freely, correcting forms not necessarily marked by the program.

The program is designed to support diverse views of language, and the different tools sometimes give incongruous feedback. One example is that the online lexicon and the inbuilt spell checker are based on different lexica, and therefore sometimes disagree as to which forms are correct. However, the students tended to seek an absolute answer. When the program could not provide this, they searched for other ways of determining the correct form, primarily with a physical lexicon or by referring to the teacher.

The lexica that the students had brought to the sessions were sometimes used to get a final answer written in black and white. They tended to view the entries in the lexicon as absolute in comparison with their own opinion, and that of the program. However, the teacher was the unquestionable authority on Swedish. After all other options of finding a correct form had been tried, the learners referred to what their teacher had said during class (excerpt 5) or, as in excerpt 6, they decided consult the written feedback from the teacher.

Excerpt 5. Ulrike and Stephanie, Session 1: Tool Interaction, Form

Ulrike: but it ((Grim)) doesn't tell // it doesn't matter // we / we / we
check later how Lena ((the teacher)) has corrected it

Excerpt 6. Maria and Jean, Session 2: Form

Handled form: "[...] bandoneón [...]"

Maria: but Erik ((the teacher)) says that we have to ask *what it's
called in Swedish* LAUGHTER

The activity was not solely about error correction. These students were interested in exploring the various features of Grim. During the first session, after attending to the

forms marked as erroneous by the grammar checker, they tried other tools:

Excerpt 7. Maria and Jean, Session 1: Tool Interaction, Form

Jean: mm infinitive phrase ((marking of infinitive phrase)) is good because there are many who / who makes mistakes with to happens ((att råkar)) for example (mhm) without saying happens ((råka)) (mhm) that can be useful

Even though the students in this case believed tools like the syntax highlighter could be useful, they scarcely used them during the sessions. The more implicit tools were tried in a playful manner and not so much for actual revision of the text. It is notable that the students' main subjects were in engineering, and they may have been interested in exploring features for the sake of investigating the technologies rather than only as a tool for writing.

4 Discussion

The previous sections have provided glimpses into the activity of these students. To answer our questions regarding tool use, we wish to discuss the observed student activity and what it tells us about what views on language and learning Grim mediates.

We have seen that the overarching interest for the students was to correct sentence-level linguistic forms. When searching for the correct form, they consulted Grim in some instances, but discussed form without the tool in other cases. This is to be expected; a more surprising result would have been if Grim were present in all or none of the language related episodes.

Likewise unsurprisingly, they were at times helped by Grim's suggestions, but were at other times led in the wrong direction. The reason is, of course, that grammar checkers are not fool-proof. This is an interesting research problem in its own right; how do we target grammar checkers specifically for learners? (E.g. Bigert, Kann, Knutsson & Sjöbergh 2004; Knutsson *et al* 2006). Here, our interest does not lie in what grammar checkers can and cannot do, but in what activity Grim stimulates.

Some of the linguistic episodes we found show that Grim does indeed stimulate discussion, and therefore reflection, concerning form. In consulting Grim, dictionaries and their teacher, students investigate problematic forms in depth. This result suggests that while false alarms pose important and challenging problems for LT, they can be used for discussion purposes in real classroom contexts. In that sense, Grim fulfils its goal of being a tool for encouraging language reflection, and it is a useful tool for Focus on Form if it is used in a context that allows and encourages means of further reflection.

It is more troublesome that some technological limitations were tackled by working around problems, and that the students accepted some faulty output as correct. The reason is that the objective for the students became not primarily to write a grammatically correct text, but to get rid of all of the error markings. Given the prominent appearance of the error checkers, this is perhaps also unsurprising, there was an urge to "tidy up" all the red markings.

After some familiarization with Grim, the students acknowledged that it did not

catch all errors. Therefore, they made some final corrections themselves, but they neither introduced new error markings (correct or incorrect), nor reverted to forms where they had questioned Grim's authority. Other tools were not extensively used, even though the students agreed that these tools might be useful precisely during the revision phase of writing. This result is perhaps more surprising; they worked towards satisfying Grim in spite of knowing better. This is no fault of the students. Instead, it shows how powerfully a CALL application can mediate views on pedagogy and language. Even when obviously wrong, its markings took precedence over the students' own judgment.

Text reconstruction and text revision entail much more than correcting forms at sentence-level, and it is quite important to get meaning across. In addition, attending to errors in individual sentences is necessary, but not sufficient. Revising a text usually entails other issues such as attending to text coherence and cross-sentence issues in form.

Encouraging students to neglect meaning (and even change it for the worse) in order to satisfy an error checker is problematic. We believe that the focus on Grim's error markings stems from an expectation that the tool will take initiative and explicitly tell students what was wrong with their texts, rather than participate in reflective activities that concern reconstruction and revision to a fuller extent. This occurs in spite of Grim being built precisely to provide several alternatives and therefore encourage reflection.

One reason for these results lies in introducing a tool that contains prominent error correction, but that is not the entire picture. In addition, the tasks we provided may be interpreted as exercises in producing correct texts, especially if students are used to being corrected when participating in similar tasks. Most language courses, including the ones our subjects attended, incorporate a large amount of corrections and guidance from the teacher. This points towards a more general problem in education, how to balance the need for guidance from a more knowledgeable source with an equally necessary need for critical thinking (in SCT, these are essential in the *Zone of Proximal Development*). Being used to error correction from humans may reinforce a view of computers as corrective tutors.

In studies similar to ours, Lindberg (2001) performed sociocultural studies of linguistic episodes in collaborative text-reconstruction, but without using computer tools. One of her findings was that collaborative dialogue is a potent means of learning language because students need to explain general linguistic patterns to their peers in order to win arguments about which form to use.

Still, it was not always the correct form that prevailed. Factors such as peer pressure, status, negotiation skills, and personality also determined the outcome. On occasion, the majority decision of the students led to a conservative approach, where they chose to write what had been read to them verbatim rather than to rely on their own confidence in alternative forms. This is similar to our informants' confidence in Grim. It suggests that the search for absolute answers is not only an effect of using Grim, but that it is also built into the task and the classroom situation.

Of course, reflecting on language via writing is a much more difficult task than correcting errors. Even for advanced, competent, and highly motivated learners it is hard to engage in a reflective conversation with their texts. Users may get linguistic

information from different tools and at different language levels at once and that can be experienced as confusing and cognitively demanding.

Our results are not merely a critical analysis of Grim, nor do we wish to rid language curricula (or CALL) of error correction. Instead, we wish to point out one of the difficulties in introducing tools such as Grim in real classroom contexts, difficulties that occur in our attempts to incorporate communicative aspects of learning languages while retaining some attention to linguistic forms.

Thus, we highlight what we believe to be a very important problem in LT-based CALL, different from the classical technological problems concerning improved linguistic analysis or “intelligence” in computer applications. Grim might be said to attempt to tackle those technical issues by providing alternative means of interaction at the level of the computer interface and the task. By doing that, we wish to bring LT-based CALL into the educational context of Focus on Form. However, when analyzing student activity, another picture emerges. The tool, in conjunction with the learning situation, leads these students to an activity where forms are indeed attended to, but the communicative parts of writing are set aside.

These issues raise awareness of the multiple layers of the task and contribute to rethinking the problem of introducing Language Technology in second language learning.

5 Conclusions

We have seen that the students’ activity can be described as one within Focus on Form pedagogy, in the sense of attending to linguistic forms while pursuing a communicative task. However, use of the tool with the tasks we provided limited the students to certain kinds of linguistic activity, primarily error correction at sentence level. The communicative tasks were transformed into activities that had more focus on errors than on expressing meaning.

Our studies suggest that there is a need to consider what view of language we wish to promote, and how that view is received by students. In our work, the CALL application was LT-based, and the tasks were based in FonF. Other kinds of CALL and other pedagogies might of course yield different results, but the question of what view of language our tools mediate nevertheless ought to be a central one.

Language Technology can be used to stimulate discussion concerning linguistic form. We believe that this discussion is beneficial for students, even if the linguistic engines on occasion provide the wrong answer. The point is to deepen reflection, which is not necessarily the same thing as providing the right answer.

Returning to the question of *mediation* done by the tool, we need to discuss not only how mature Language Technology is with respect to how precise it is in reporting errors, but also what it tells students about language. If the linguistic engines are built to detect errors, they may reinforce a view of language as rule-governed object and on pedagogy as mainly corrective, even if alternatives are provided.

The linguistic engines as presented to learners affect which tools they will use. Error reporting is very attractive for students, and is in many cases used for reflection. However, its prominent reports lure students into relying on it to take initiative and tell them what to do with their texts, rather than reflect on what they are trying to express.

Furthermore, people always form and reform attitudes and beliefs towards the tools they use. In our case, a tool for reflecting on language was transformed into a tutor that corrected students. It is perhaps not too far-fetched to believe that other LT-based tools may also be expected to provide “correct answers” in a similar way, because this is what computers in education traditionally do (unless used in a purely communicative sense)?

Finally, the task and the educational environment in which it takes place affect the activity. Learners often wish to know the correct answer to their questions. This is of course quite reasonable. Learners need expert advice, and should expect to receive it. Still, in order to progress in their new language they also need to participate in more and more demanding tasks, some of which have no single “correct” answer, such as text-reconstruction and full text revision.

In order to develop and investigate further innovations in LT-based CALL, it is important to consider its role in this kind of task as well as traditional ones. Expanding on this point, it is quite important to consider what view of language any particular CALL tool mediates. In this respect, sociocultural theory is a way of highlighting what the tools we build really tell our students.

Acknowledgements

The research reported here was partially funded by Vetenskapsrådet (The Swedish Research Council) as part of the project “The use of language tools for writers in the context of learning Swedish as a second language”, and by Nordforsk. We are grateful to the students who allowed us to record their interactions. We also wish to thank professor Inger Lindberg and the seminar at The Institute for Swedish as a Second Language, Göteborg University, for fruitful discussion, as well as our anonymous reviewers for their insightful comments. The Open Source software Transana (Fassnacht & Woods, 2006) created by Chris Fassnacht and currently developed by David K. Woods greatly facilitated our analysis work.

References

- Bigert, J., Kann, V., Knutsson, O. and Sjöbergh, J. (2004) Grammar checking for Swedish second language learners. In: Henrichsen, P. J. (ed.), *CALL for the Nordic Languages*, Copenhagen Studies in Language, **30**. Samfundslitteratur, 33–47.
- Borin, L. (2002) What have you done for me lately? The fickle alignment of NLP and CALL. Presentation at the EUROCALL 2002 pre-conference workshop on *NLP in CALL*, August 14, 2002, Jyväskylä, Finland.
- CALICO ICALI, <https://calico.org/p-10-Special%20Interest%20Groups.html#icali>
- Chapelle, C. A. (1997) Call in the year 2000: Still in search of research paradigms? *Language Learning & Technology*, **1** (1): 19–43.
- Chapelle, C. A. (1999) Research Question for a CALL research agenda: A reply to Rafael Salaberry, *Language Learning & Technology*, **3** (1): 108–113.
- Chapelle, C. A. (2001) *Computer Applications in Second Language Acquisition*. Cambridge University Press.
- Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Dewey, J. (1991) *How we Think*, New York: Prometheus Books.

- DiCamilla, F. J. and Lantolf, J. P. (1994) The linguistic analysis of private writing *Language Sciences* **16** (3/4): 347–69 .
- Doughty, C. and Williams J. (eds.) (1998) Focus on Form in Classroom *Second Language Acquisition*. Cambridge: Cambridge University Press.
- EUROCALL NLP SIG, <http://www.eurocall-languages.org/sigs/langproc.html>
- Fassnacht, C. and Woods, D. (2006) Transana v2.12. <http://www.transana.org>. Madison, WI: The Board of Regents of the University of Wisconsin System.
- Gánem Gutiérrez, G. A. (2006) Sociocultural theory and its application to CALL: A study of the computer and its relevance as a mediational tool in the process of collaborative activity. *ReCALL*, **18** (2): 230–247.
- Grim, available for download <http://skrutten.nada.kth.se/grim/index-en.html>
- Haas, C. (1996) *Writing technology. Studies on the materiality of literacy*, New Jersey: Lawrence Erlbaum Associates.
- Haas, C. (1999) On the relationship between old and new technologies. *Computers and Composition* **16** (2): 209–228.
- Knutsson, O., Cerratto-Pargman, T., Severinson-Eklundh K. and Westlund, S. (2006) Designing and developing a language environment for second language writers. *Computers and Education*. In press.
- Krashen, S. (1985) *The Input Hypothesis: Issues and implications*. Pergamon.
- Lantolf, J. P. (2000) Introducing sociocultural theory. In: Lantolf, J. P. (ed.), *Sociocultural Theory and Second Language Learning*. Oxford: Oxford University Press, 1–26.
- Lantolf, J. P. (2001) (S)econd (L)anguage (A)ctivity theory: Understanding second language learners as people. In: Breen M. (ed.) *Learner Contributions to Language Learning: New directions in research*. London: Longman, 141–159.
- Lantolf, J. P. and Thorne S. L. (2006) *Sociocultural Theory and the genesis of second language development*. Oxford: Oxford University Press.
- Lindberg, I. (2001) With a little help from my friends: On the role of collaborative tasks in second language learning. In: Fraurud, I. K. and Hyltenstam, K. (eds.), *Proceedings from the 8th Nordic Conference on Bilingualism*, Stockholm: Rinkeby Institute for Multilingual Research and Centre for Research on Bilingualism.
- Long, M. (1991) Focus on form: A design feature in language teaching methodology. In: de Bot , K., Kramsch C. and Ginsberg, R. B. (eds) *Foreign Language Research in Cross-Cultural Perspective*. Amsterdam/Philadelphia: John Benjamins, 39–52.
- Long, M. H. and Robinson, P. (1998) Focus on form: Theory, research and practice. In: Doughty, C. and Williams, J. (eds.), *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press, 15–41.
- Nerbonne, J. (2003) Natural language processing in computer-assisted language learning. In: Mitkov, R. (ed.) *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press: 670–698.
- Olson, D. R. (1995) Writing and the mind. In: Wertsch, J. V., del Rio, P. and Alvarez, A.(eds.), *Sociocultural studies of mind*. Cambridge: Cambridge University Press, 95–123.
- Salaberry, R. (1999) CALL in the year 2000: Still developing the research agenda. *Language Learning & Technology*, **3** (1): 104-107.
- Swain, M. (2000) The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In: Lantolf, J. P. (ed.), *op. cit.*: 97–114.
- Transana analysis software <http://www.transana.org/>
- Vernon, A. (2000) Computerized Grammar Checkers 2000: Capabilities, limitations, and pedagogical possibilities. *Computers and Composition*, **17**: 329–349.
- Vygotsky, L. (1962) *Thought and language*. Cambridge, MA: MIT Press,.
- Vygosky, L. (1978) *Mind in Society: The development of higher psychological processes*.

Harvard: Harvard University Press.

Wertsch, J. V. (1991) *Voices of the mind. A sociocultural approach to mediated action*. Harvard: Harvard University Press.

Wertsch, J. V. (1995) The need for action in sociocultural research. In: Wertsch, J. V., del Rio P. and Alvarez, A.(eds.), *Sociocultural studies of mind*, Cambridge University Press, 56–74.