

Analyzing Discourse Synthesis – Use of Semantic Information for Collaborative Writing in Secondary Education

Crina Damşa, InterMedia, University of Oslo, Postboks 1161 Blindern 0318 Oslo, Norway,
crina.damsa@intermedia.uio.no

Gijsbert Erkens, Jeroen Janssen, Research Centre Learning in Interaction, Utrecht University, the Netherlands,
Email: G.Erkens@uu.nl, J.J.H.M.Janssen@uu.nl

Abstract: Discourse synthesis refers to writing activities that involve synthesizing information from multiple textual sources for writing a new text. This study investigates the way groups of secondary education students use semantic information from multiple sources, while undertaking a collaborative, computer-supported writing history task. We analyzed the products of 10 groups (32 students) and the chat discussions of the group members, and we assessed the quality of the group products. The results of the study offer an insight into how students select and integrate semantic information. Most noticeable results show that groups synthesized the information selected from sources based on ideas rather than sources and created new conceptual connections among these ideas. Relationships were found between the quantity and type of information and the products quality, while weak or no relationships were found between type and frequency of chat talk with the degree of integration in the group products.

Introduction

Discourse synthesis refers to reading and writing activities within a specific type of writing task, which requires subjects to synthesize information from multiple textual sources (see Spivey, 1995). To a certain extent, discourse synthesis is employed at all educational levels, from primary school to university and in all disciplines. The products of this activity can vary from reports to essays or literature studies. Such tasks are often used for evaluating students' capacity to synthesize information from different sources and to build an argument. The educational practice shows that the evaluation of this type of written tasks is mainly realized by only considering the final text, and very rarely based on the process that took place in order to produce that text. Nevertheless, there is more about discourse synthesis tasks than this final text. The cognitive and meta-cognitive processes taking place while performing such tasks are of importance. Research studies show that a series of cognitive activities differentiate this type of task from other writing tasks and often make the difference between a good and a mediocre composition (Segev-Miller, 1997). While many studies of discourse synthesis focus more on the product and its characteristics (content, composition), the more recent ones attempt to gain insight into the cognitive and meta-cognitive landscape of the discourse synthesis process. This research study focuses on one specific aspect of the discourse synthesis process, i.e., the way students use the semantic information from the multiple textual sources.

Discourse Synthesis

Studies on discourse synthesis tasks have been conducted with two purposes: firstly, to identify the cognitive and meta-cognitive mechanisms of the process and secondly, to identify the differences in performance between successful and unsuccessful synthesizers, and the factors that might cause these differences. However, strategies identified by the mentioned studies are not unique to the discourse synthesis process. Many strategies identified in the synthesizing process have been found also in general writing task, such as: selecting (Kantz, 1998; McGinley, 1992; Higgins, 1992; Spivey & King, 1989), planning (Kantz, 1989, McGinley, 1992); elaborating (Higgins, 1992), organizing (Higgins, 1992; McGinley, 1992; Kantz, 1989); evaluating, (Higgins, 1992; Segev-Miller, in press), revising (Kantz, 1989; Higgins, 1992; McGinley, 1992).

Some analyses also provided an insight into the characteristics of the discourse synthesis tasks. A number of studies concerned with the cognition within the writing tasks pointed at the fact that synthesizing information from sources occurs both at textual and conceptual level. Spivey (1997) calls the cognitive processes in discourse synthesis processes *textual transformations*. The production of the new text happens by cognitive means which are: *selecting, organizing and connecting transformations*. Segev-Miller (1997; in press) tried to gain a more comprehensive view on the cognitive processes underlying the performance of this task at academic level. The element of novelty in Segev-Miller's approach is the category of *transforming or intertextual strategies*. These can be briefly described as cognitive strategies that subjects use in order to synthesize information provided by more textual sources and use it for contracting a new idea, phrase or text. Based on her studies, Segev-Miller (1997) categorizes transformations in *conceptual, rhetorical and linguistic*.

We explore first activities at the conceptual level. The activity of *selecting* information occurs mainly during reading, in the context of meaning making. Spivey (1997) maintains that, during reading, meaning is constructed and then text is produced in order to 'signal' that meaning to others. She considers it important to

distinguish between two products of discourse processes: mental products and textual products. The mental product is actually the meaning the individual is constructing using the reading text together with prior knowledge as input. The writer, actively producing the mental meaning, is also producing a written product, meant to be read by the others. According to Spivey and King (1989), selecting is guided by a very clear goal that the reader sets, which is to find relevant input for the own text. It is not possible or necessary for the reader to use all information from the given sources. A number of criteria are identified based on which readers select information (Spivey, 1995). These criteria are: *textual* relevance, which refers to the hierarchical placement of content within the whole text; *intertextual* relevance refers to the information occurring in more than one source and again, the placement of the content within the various text sources; the *contextual* criterion involves that writers choose the information that is most relevant to the type of message being written; the *rhetorical relevance* refers to the fact that writers use certain information from sources for particular audiences, and adapt their selection process to the type of audience.

Although named differently, comparable conceptual activities have been identified by Spivey and Segev-Miller in their studies. Spivey (1997) identifies more types of connecting transformations: the *topical connections* are made within texts as well as across texts and have the purpose to maintain the coherence of the discourse clause by clause; the *whole-text connections* are transformations that occur at a larger scale, and are usually intertextual; *nonlinear* connections are comparable with the connections made while using hypertext. Segev-Miller (1997) uses the term *conceptual transforming*, which consists of deliberate intertextual processing of the source texts in order to identify conceptual connections between them, and the creation of macropropositions, in order to connect the proposition from the sources. When exploring activities at *rhetorical* level, Higgins (1992) identifies activities such as structuring, shaping and re-shaping materials from the various source texts. The connection between the separate semantic entities is realized by means of manipulating propositions originating from the source, such as arranging text into high- and low level propositions, discovering relations between ideas, or looking for super-ordinate categories in order to subsume items. Spivey and King (1989) consider organizing as supplying the content with a new structure. Writers are busy creating different semantic chunks which are fit into the organizational patterns of the original text, but mostly are being used to generate a new kind of organizational pattern (Spivey, 1995). In this type of organizational strategy the writer provided an integrated presentation of the material that has been selected from different sources. Segev-Miller (1997) identified rhetorical activities, which serve to translate the conceptual connections to text. Five rhetorical strategies were identified and classified, based on the degree of cognitive demands every strategy requires. *Summarizing one source text* is a less demanding strategy. *Listing* the source text, compared with Bereiters and Scardamalia's (1987) knowledge-telling, is a strategy which consists of summarizing every source text and presenting these summaries after each other. *Incorporating source texts in one source text* is applied by using one text as a frame for incorporating information; and by adding information from other texts order to replace deleted information from the frame text. The strategy of *decomposing and recomposing* source texts is considered as being of intermediate difficulty, and consists of 'synthesizing by ideas rather than by authors' (Segev-Miller, 1997, p. 22). It involves breaking one sources text into propositions, and then looking across the other source texts for propositions that can be related to these propositions from the first source. Finally, *synthesizing the source texts* involves more than two source texts, requires a high level of intertextual processing and is considered the most difficult rhetorical activity.

Finally, Segev-Miller (1997) identifies *linguistic strategies*, used in order to compensate for the low degree or lack of conceptual transforming and intertextual processing. Examples of such activities are *speech acts*, which consists of using certain type of semantic means (usually words) in order to create the impression of connection between the different ideas that are presented; or *lexical repetition*, such as paraphrasing or repetition, used to increase the degree of cohesion at thematic level in their texts.

A Model of Writing Based on Multiple Sources

We constructed a model of text writing, which attempts to emphasize the way information from multiple sources is used for writing a new text. A first phase in the process includes *the selection of relevant information*. In previous studies, little was said about the quantity and type of the selected information, or the choice for certain sources during selection. *Integration of information* is another phase. In this phase, subjects try to create a conceptually meaningful representation based on the selected information, in order to compose the new text. In addition, the research studies mentioned activities at meta-cognitive level, such as planning, monitoring and evaluating, which are meant to offer the guidance and support for the actual writing in process. The recurring assertion of researchers regarding the cognitive activities in writing is that this process of writing is not just a rigid sequence of processes and sub-processes, but more a set of activities organized by the writers in a certain order, depending upon intermediate results of the process. The result of these activities is visible in the intermediate versions of the text (drafts). The way students deal with information from sources during these phases, the type of information they use from sources and how they manage their writing process will be reflected in the final product.

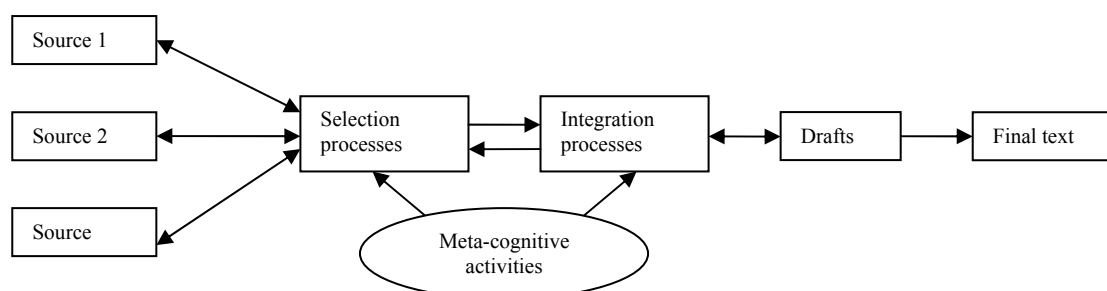


Figure 1. Model of Writing from Multiple Sources.

In relation to the above, the following research questions will be answered: 1. *How do groups of secondary school students make use of semantic information from multiple sources while undertaking a computer supported inquiry history task?* and 2. *What meta-cognitive activities are reflected in chat discussions of groups of secondary school students, when undertaking a computer supported discourse synthesis history task?* We expect the ‘quality’ of written text to correlate with the amount of information that students use from the given sources; in other words, students that select and integrate more information from sources will write better texts. Another expectation is that the quality of written text would correlate with the type of information that students use from sources; in other words, students that use information from sources that are considered essential will write better texts. In addition, we expect that there will be a relationship between the type (on- or off-task) and frequency of chat talk with the degree of integration in the final text product.

Methods

The participants in this research project were 32 (mean age of 16 years, $SD = .58$) eleventh-grade students, along with their history teachers, from a secondary school in the Netherlands. Participating students were randomly assigned to a group of three or four by the researchers, in order to obtain heterogeneous groups with respect to gender and ability. During the experiment, students collaborated in the *Virtual Collaborative Research Institute (VCRI)* designed by Erkens, Janssen, Broeken & Jaspers (2002). VCRI is a groupware program designed to support groups of students during collaborative projects or inquiry tasks. A number of tools, such as *Chat*, *Sources* and *Co-writer*, were used in order to perform this collaborative writing task, and other tools e.g., the *Statusbar*, *Forum*, *Planner*, were employed to organize the process.

Students worked in collaboration on a history inquiry task with the title ‘Witches and the prosecution of witches’. Based on the six given information sources (and on additional sources, found independently) students had to perform seven assignments, by answering different questions with regard to the topic and write their answers as short descriptive texts, or argumentative essays. Students worked together in groups of 3 or 4, using the VCRI environment to collaborate. In VCRI, groups had access to the sources, could make notes, drafts, communicate with the other members of the group, etc. Through the Co-Writer every member of the group could see how the draft (shared text) evolved to a final version, and the other’s contribution to the text. The result of their work was an argumentative text of 700 words about witch prosecution. The sources used for this assignment were original historical texts or translations of original texts about witch prosecution in the 16th and 17th century. The student groups worked the inquiry group task for a period of four weeks. In the first lesson, the teachers introduced the task and gave the necessary instruction regarding to the task, and the most important features of the CSCL-environment were explained. Seven more history lessons were devoted to the inquiry group task. Students were allowed to work on the inquiry task outside normal classrooms hours.

In order to answer the first research question, two types of analyses were conducted. First, we conducted a *semantic text analysis* of the text sources, of the drafts and of the final group products. For this purpose, a parsing procedure was developed based on the theory of semantic representation of Kintsch and Van Dijk (1978). Two types of parsing rules were applied: the macro rules defined by the semantic representation theory (generalization and construction), and the second set of rhetorical rules, developed in order to determine the morphological structure of the propositions. A first aspect treated by these rules is the structure of the (semantic) propositions and the use of grammatical aspects. According to these rules the standard structure of a proposition is: *subject, verb, object, specification of the object*. The subject and the verb are compulsory elements of the unit. Some examples of these rhetorical rules are: the nouns are always singular, with the exception of the situation when a generalization is necessary; or verbs are always used with their infinitive form, present tense and passive voice. This is an example of text: “*Amongst the women was one of unbelievable beauty who was telling all kinds of fabulous stories about what she had done; amongst others, she told she could fly...*” (Translation from Dutch from *Batavische arcadia*, by Johan van Heemskerck), which can be parsed as follows: *Woman is beautiful, Woman says (woman can fly)*. We constructed two different lists of words and their synonyms, used on sources or used in the text assignments. One of the synonyms was chosen and consistently used in all situations. Examples are: devil, was used for *devil, spirits, demons*; illusion(s) was used

for *hallucination, visions*; women was used for *girls, women*. The second list regarded the syntactical components of the propositions. Examples from this list (verbs list): accept was used for *to confess, to accept*; accuse was used for *to officially accuse, to call for blame, to give the blame*.

Once sources and group products were parsed, we analyzed the way groups selected and integrated this semantic material. The following aspects were considered: 1) The *quantity of semantic information* selected from sources and used in the drafts and final group products; 2) The *type of semantic information* - from core of secondary sources; 3) The *origin of the selected information* (from which source). When analyzing the *integration of semantic information* from sources in the text products three dimensions have been considered: 1) The way in which selected *semantic information evolved along the drafts*; 2) In which *succession* students arrange the *semantic information selected from sources*; 3) The *degree of integration* of semantic information selected from the given sources in the final products. The number of propositions was chosen as an indicator when quantifying the semantic information. A proposition was considered the equivalent of a proposition that carries semantic meaning independently, or of a construction of two propositions that need each others' completion in order to carry semantic meaning. For analyzing the *degree of integration of semantic information selected from the given sources* in the final products, the number of switches made between propositions selected from one particular source to propositions selected from another given source was used as indicator. The following types of switches were examined: between segments of propositions selected from the given sources (Sbb); between units selected from the same given source (Sb1b1); between a unit selected from a given source and a unit selected from another given source (Sb1b2). For examining the *degree of integration* between propositions selected from the given sources and propositions selected from other sources (others text books, or written sources found by students themselves, internet sources), the following switches were examined: between propositions selected from a given source and propositions selected from another source, no matter in which direction (SB0); between propositions selected from a given source and propositions selected from another source (Sb0); between propositions selected from another source and propositions selected from given sources (S0b); between propositions selected exclusively from other sources (S00).

The second type of analysis was *the assessment of quality of group products* and was conducted by using an assessment instrument. This instrument focused on three major aspects of the group products: use of sources, content and/of argumentation, and structure and language use. Every assessed aspect was rated on a three point scale (from 0 to 2). *Use of sources*, consisted of two elements: complete use of sources and copying and pasting. *Content and argumentation* considered what students actually did with the information from the given sources (or from other sources) in order to compose a coherent text. The component *Structure and language use*, focused on the organization of the text and on language use. An 80% total agreement between the two evaluators was reached when using this instrument, with a value of .66 for the Cohens kappa coefficient.

In order to answer the second research question, an *analysis of chat discussion protocols* was conducted. Based on iterative readings of the chat protocol a coding system was developed, consisting of the following coding categories: 1) Selection - statements about the selection of semantic information from the given sources (e.g., "...and for con-arguments that witches are blamed for everything"); 2) Integration - statements regarding the structure of the group products and the use of the selected semantic information in the texts (e.g. "...maybe we should say that witch prosecution was good, because if a witch is prosecuted other witches get to scared to go on with witchcraft..."); 3) Planning - statements regarding the planning of selection, integration and writing acts (e.g. "who is going to do the comparison?"); 4) Monitoring - statements regarding the monitoring of selection, integration and writing acts (e.g. "I begin with the arguments for assignment 3"); 5) Evaluation - statements regarding the evaluation of selection, integration, and the progress of the text (e.g. "no, there are too many arguments for witch prosecution"); 6) General Information - general statements on the progress of the group product and the writing process (e.g. "I'll look at some arguments from supporters, will you look at some from the opponents?"); 7) Other - any other statement not included in the other categories (e.g. "I will go on with chapter 3"). To select and code the text material we used the *Multiple Episode Protocol Analysis* (MEPA), a software program (Erkens, 2003) which offers the possibility to visualize and categorize the VCRI files.

Findings on Use of Semantic Information

Regarding the selection of semantic information we report on a number of aspects. Concerning the *quantity of semantic information selected from sources*, percentages of the propositions selected from sources show how quantity of information is distributed per draft, per group along the drafts and in the final product. The average quantity of semantic information selected from sources represents more than a third (35.9%), from the semantic information identified in the final products. A noticeable aspect is the high range between different groups. In both drafts and final texts the differences between groups in the selected number of propositions were big: in draft 1 the range is 79%; in draft 2 it is 65.9% and in the final text it is 58.5%. These differences show that some groups relied almost completely on the given sources for collecting semantic information for their assignment. Other groups, such as group 201 (9.9%), used especially information selected from other than the given sources.

Regarding the *type of semantic information from sources* on average 35.9% of the semantic content in the final group products is represented by information selected from the given sources. More than half (55.5%) of the selected and used propositions was core information; about a third (33.5%) was secondary information. These results show that groups concentrated on selecting *core information* and that students were less interested in selecting semantic information unrelated to the assignment. The *quantity of semantic information selected from each source* was quantified in relation to the total number of propositions per group product. The pie chart in Figure 2 shows that almost two thirds (63.8%) of the quantity of semantic information used for writing the final group products was selected from other sources than the six given sources.

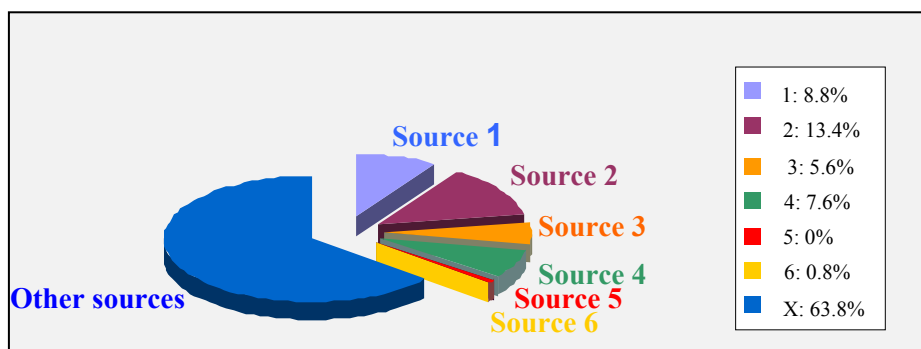


Figure 2. Quantity of Semantic Information Selected from Each Source and from Other Sources.

Only 35.2% of the propositions are selected from the given sources and the chart shows the division of the quantity selected from each source, in relation to the total number of propositions identified in the final group products. These proportions also indicate that groups had a strong preference for certain sources; especially for source 2 (13.4% from the total number of selected propositions). The content of the sources explains this preference: this source contains a great number of explicit pro- and con-arguments that could almost literally be used when writing the assignment text. Source 1 (8.8%), source 3 (5.6%) and source 4 (7.6%) contain a number of clear arguments as well. From these sources content was selected, but in a lower percentages. No units were selected from source 5. The percentages of propositions selected from other sources or invented by groups themselves are unexpectedly high in comparison with the percentages of units selected from the given sources.

When examining how students *integrated semantic information* selected from the given sources in their texts, a number of aspects were considered. The average quantity of semantic information selected from sources represents more than a third (35,9%) from the semantic information identified in the final group products; in the second draft, the percentage was four percent higher than in the final version (39.4%); and in the first draft half (50.3%) of the semantic content was selected from the given sources. It is noticeable that group 201 is the only group for which the percentage of the propositions selected from sources is higher in the final group product than the proportion in the first and second draft. For all the other groups this proportion is higher in the first draft and gradually becomes smaller, while the writing activities process heads towards the end. Regarding the *succession of the propositions selected from sources in the new text*, no stable succession pattern could be determined when looking at the organization of the units in each group product and when comparing the succession of the sources between the group products. Even so, some features of the succession could be named: six out of the ten groups started their text with propositions selected from source 2; in eight out the ten products source 2 was used in the first half of the texts; with one exception, sources 1 and 4 were used in the second half of the text; information units from source 3 and source 6 were randomly distributed in the semantic structure of the texts; information units from sources 1, 2, 3, and 4 were used on various locations in the texts.

To determine the *degree of integration*, two aspects were examined: the integration of only propositions selected from given sources and the integration of all propositions (from given and other sources). From the switches made between propositions selected exclusively from the given sources an average of 87.7 % were made between propositions selected from the same source; 12.3% were made between propositions selected from different given sources. This last percentage, together with the percentage indicating the integration of the segments of units, shows that the degree of integration of information units selected from different given sources was rather poor. Figure 3 presents the degree of integration per group in the form of switches between: units selected from given sources (Sbb, blue bar); units selected from given sources and units from non-given sources (SB0, purple bar); units from other sources (S00, white bar). With only one exception (group 203), the percentages of switches between units selected from given sources and the units from other (SB0) is higher than the percentage of switches between units from other sources (S00). Also, in five of the ten texts the percentage of switches between units from given sources and units from other sources (SB0) is lower than the units only selected from given sources (Sbb).

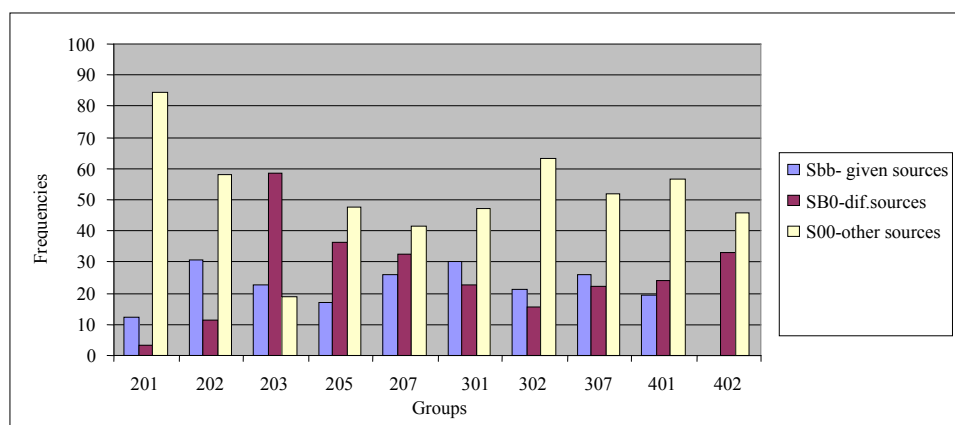


Figure 3. Percentage of Switches per Group.

Overall, the results show that 25.6% of the switches were made between units from different sources. Furthermore, 22.9% of the switches were made between units from the same source. Half of the total switches were made between units selected from other sources, and in this case there is no specification whether these propositions were selected from. The low percentage of switches between units selected from sources and units from other sources and the high percentage of switches between other sources supports the statement that the degree of integration is moderate.

A correlation analysis conducted to determine *the relation between quality of group products and the quantity and type of semantic information* shows a positive, but low ($r = .24$, $p < .10$) and non significant correlation between the obtained scores and the quantity of semantic information selected from sources. This is a correlation in the expected direction but does not confirm the expectation formulated at the beginning of this study, that groups which will select more semantic information from the given sources will also score better by the assessment of their final products. The result reveals that the obtained scores were influenced by the quantity of semantic information selected from sources, but not in a decisive manner. On the other hand, the correlation analysis *between the assessment scores and the proportion of core information selected from sources* showed a much higher degree of correlation. A positive correlation of .69 was found, and this correlation was also statistically significant ($p < .05$). The expectation that the groups that selected more semantic information labelled as core information will receive higher scores is confirmed by this result.

Findings on Meta-Cognitive Activities and Assessment of Group Products

The percentages calculated in relation to the total number of chat-statements indicate that statements about *Planning* occurred least frequently (0.4 %) and the statement on *Other aspects* most frequently (24.1%). The *Selecting* and *Integrating activities* (defining for the discourse synthesis activities) are discussed in 1.1%, respectively in 2% of the total statements. Also the discussion of *Evaluation* and *Monitoring* activities took place infrequently. In relation to the coded statements, the results show that more than half of the statements (58%) belonged to the category *Other aspects*. Statements coded as *General information* represented a third of the total. Planning, monitoring and evaluating were discussed in 9%, while selecting and integrating activities were only discussed in 7.6% of the coded statements. Furthermore, the statements regarding coded activities represented almost half percent (43.7%) of the total number of statements.

These results indicate that little discussion was held between the group members during the writing of the assignment. Observations indicated that group members split tasks and performed them without much consultation. The results show, in general, that the category *Other aspects* is strongly represented in the discussions of all groups. The discussion of one group (307) represents an exception from these findings. This group discussed about selection and integration of information from sources in their text. A qualitative analysis of this group's discussion showed that group members selected sequences of semantic information from sources individually and then discussed it with the other members, permanently discussed about the integration of selected information in the new text, and that the content, order of ideas, structure of the text, and conclusion were decided in consultation. Monitoring activities took place and, close to the end of the process, every group member performed a last evaluation of the text and discussed this with the group.

The relationship between *assessment scores* and *chat discussion categories* is reflected by a series of correlations. The correlations between the results and the categories *Selecting*, *Integrating*, *Evaluating*, *General information* and *Other* are positive, but low. The correlations between results and the categories *planning* and *monitoring* are negative. None of the correlations are statistically significant.

Table 1: Correlations between chat categories and a) assessment score and b) degree of integration.

Category	Correlations	
	Assessment scores	Degree of integration
Selecting	.28	-.20
Integrating	.32	-.19
Planning	-.36	-.41
Monitoring	-.13	.19
Evaluating	.13	-.12
General info	.22	-.04
Other	.26	-.31

These correlations show that discussing the selection and integration of information (from given or not-given sources) guaranteed a certain quality of the written text, and that also the fact that these activities took place is reflected in the better results obtained for assignment 3. As the results presented in Table 1 show, the correlations between the degree of integration of semantic information selected from sources and the meta-cognitive activities are, with one exception, negative. None of the correlations are statistically significant. These negative correlations indicate that there was no relationship between the frequency with which groups discussed about discourse synthesis activities and the degree of integration of propositions.

Summary of Findings, Discussion and Conclusion

In the present study we investigated the way groups of students use semantic information from multiple textual sources for writing their own text in collaboration. The first research question focused on how groups of secondary school students select semantic information from multiple sources when performing an inquiry task for the discipline history. When analyzing the quantity of semantic information selected from sources, the conclusion is that the differences in the quantity of selected information between groups were considerable. When distinguishing between the categories of semantic information offered by the given sources (core, secondary and other information), a selecting pattern was identified. All groups selected a higher quantity of core information, merely consisting of arguments pro- and counter witch prosecution. This category of semantic information was important for the quality of the final group products, as shown by the positive correlation between the assessment results and the categories of selected information. A much lower quantity of secondary semantic information was selected and used in the new text, and hardly any from the category of *Other information*. An aspect that deserves attention is the work strategy the majority of the groups used when performing this writing task. Firstly, semantic information from the given sources was selected. This information was organized based on contextual criteria (superordinate semantic categories), and not based on the source succession. The propositions selected from sources were organized in semantic sequences, most of the time units from the same source together. Thereafter, more sources of information were sought and new semantic information was selected from these sources. This information was also organized in semantic sequences, but it occurred more often that semantic information from these sources was intertwined with semantic information selected from the given sources. This semantic structure was completed by groups' own constructed propositions, which fulfilled the role of connecting propositions between semantic sequences. Regarding the degree to which students integrated semantic information from sources, the results showed that the degree of integration of semantic information from the given sources was situated at a quarter of the maximum degree possible. The result of such a structuring strategy is a lower degree of integration. The expectation that groups which use more information from sources will write better texts was not confirmed by the findings. The expectation that the groups that selected more core information will receive higher scores was confirmed.

The second research question concerned *meta-cognitive* activities initiated by students in order to regulate their information selecting and integrating process. The assumption was that statements concerning planning, monitoring, and evaluating the writing activities would be identified in the discussions students held through the chat. The results showed little presence of discussions regarding selecting and integrating semantic information from the sources. Additionally, other meta-cognitive activities which are characteristic for writing tasks in general, such as planning, monitoring or evaluating occurred rarely in groups' discussions. The content of the virtual discussions was dominated by statements regarding general aspects in relation with the task content and performance. The reason for this moderate representation of the meta-cognitive activities can be found in the fact that most of the groups chose to distribute tasks amongst the members of the group, each of them being responsible for one or more assignments. The expectation that there will be a relationship between the type (on- or off-task) and frequency of chat talk with the degree of integration in the final text product was not confirmed.

The characteristic of selecting information from sources based on the type of information these sources offer can be discussed in relation to the way students integrated the selected semantic information in the new text. The overviews of the semantic structure of the new texts show that groups selected semantic information from sources and organized it according to a content-based structure. According to the classification of

rhetorical transforming made by Segev-Miller (1997), the groups synthesized the information selected from sources based on ideas rather than based on authors (sources). The strategy adopted by these groups can be classified as *decomposing and recomposing* the source texts, of breaking the source into propositions and looking across other text(s) for related propositions, which can also be placed under the same superordinate semantic categories.

A number of issues need to be considered in relation to this study. First, performing a task in a group context had certain consequences for the way a student approached the task and the responsibilities to be taken when performing a writing task. Moreover, when analyzing the products of a collaborative writing task the risk appears that the analysis of discourse synthesis activities performed at individual level would not be detailed enough. In this research project, the contribution of the individual participants is studied in the context of collaborative group work, which might overshadow some of the details of the individual input. On the other hand, the analysis of the writing activities when performed in a group context can offer extra insights into the mechanisms of the process. The necessity of consultation and agreement between the group members results in explication of individual ideas and stances, which provide valuable information about the collaborative process. Alas, in this research this potential advantage could not be used because of the reduced amount of explicit discussion on discourse synthesis activities. Second, the task used in this study was an authentic task used in the Dutch history curriculum, and not created especially for this investigation on discourse synthesis activities. Therefore, the sources were not selected with an eye on the discourse synthesis task. The texts were very complex and presented a high degree of difficulty. Especially the selection of information might have been affected by this complexity. It is possible that students avoided selecting information from the difficult sources and relied more on the rather accessible source texts. Third, an inter-rater reliability analysis should have been applied for establishing the reliability of the coding system for the chat protocols. Due to practical reasons this analysis was replaced by the researcher scoring the analyzed dialogues twice.

In sum, this study provided some valuable information on the way groups of students select and integrate semantic information from multiple textual sources, and the discussions of the groups while performing this task. Further research is needed for a more specific investigation of the degree of integration of semantic information from various sources, wherein the use of information from other than the given sources or students' prior knowledge are also considered. Moreover, further investigation of the group discussions around the task-related activities is recommended.

References

- Erkens, G., Janssen, J., Broeken, M., Jaspers, J. (2002) *VCRI*. Retrieved March 23 from Utrecht University, department of Educational Sciences Web site: <http://edugate.fss.uu.nl/~croci/vcri.html>.
- Erkens, G. (2003). Multiple episode protocol analysis (MEPA). Version 4.9. Utrecht University, The Netherlands.
- Higgins, L. (1992). Reading to argue: helping students transform source texts. In A. Penrose & B. Sitko (eds.), *Hearing ourselves think: cognitive research in the college writing classrooms*. (pp. 77-101). New York: Oxford University Press.
- Kantz, M. (1989). *Written rhetorical synthesis: processes and products*. (Technical report no. 1). The Center for the Study of Writing, Pittsburgh, PA & Berkeley, CA, Retrieved on August 15, 2005, from National Writing Project, U.S. Department of Education web site: <http://www.writingproject.org/TR17.pdf>.
- Kintsch, W. & Van Dijk, T. (1978). Towards a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- McGinley, W.J. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly*, 27 (3), 226-249.
- Segev-Miller, R. (in press). *Cognitive processes in discourse synthesis. On the differences between successful and unsuccessful synthesizers*. A paper presented at the 9th International Conference on the EARLI on Writing, Geneva, Zwitserland (September, 2002).
- Segev-Miller, R. (1997). Cognitive processes in discourse synthesis: The case of intertextual processing strategies. In *Cognitive processes in discourse synthesis*. (Dissertation). The Hebrew University of Jerusalem, Retrieved on Mei 15 2005 from <http://www.staffs.ac.uk/personal/sciences/mt15/writing.doc>.
- Segev-Miller (2004). Writing from sources: the effect of explicit instruction on college students' processes and products, in *Educational Studies in language and literature*, 4, 5-33.
- Spivey, N.N., & King, J.R. (1989). Readers and writers composing from sources. M.R. Ruddell (ed.). In *Theoretical models and processes of reading*, (pp. 668-694). Newark: International Reading Association.
- Spivey, N.N. (1995). Written discourse: a constructivist perspective. In Steffe, L.P. & Gale, J. (1995). *Constructivism in education*. Hillsdale: Lawrence Erlbaum Associates.
- Spivey, N.N. (1997). *The constructivist metaphor. Reading, writing and the making of meaning*. San Diego: Academic Press.