Cognitive Models of Writing

David Galbraith, Stoke-on-Trent

ISSN 1470 – 9570
Cognitive Models of Writing

David Galbraith, Stoke-on-Trent

Abstract

This paper reviews models of the cognitive processes involved in writing. It sketches their development from an early emphasis on the thinking behind the text and the impact of cognitive overload on this, through more dynamic models emphasizing the interaction between thinking and text production processes, to more recent models emphasizing the constitutive role of text production in the development of the writer’s thought. In the course of the review it considers the implications of these models for research on writing in L2. These include a consideration of (i) varying goals and genres in L1 and L2, (ii) the impact of linguistic fluency in L2 on higher level thinking processes, (iii) strategies for managing the writing process in L2, (iv) the maintenance of conceptual representations during text production in L2, and (v) the impact of L2 writing on the development of the writer’s understanding during text production.

1. Introduction

Two themes have dominated psychological theories about the cognitive processes involved in writing since their inception in the early eighties. The first is the basic insight that writing is not simply a matter of translating preconceived ideas into text, but also involves creating content and tailoring the way this is presented to the needs of the reader. Writing is as much a matter of discovering or inventing the thought to be expressed in the text as it is a matter of expressing it in an appropriate and convincing way (Flower & Hayes 1980a). The second is that, because writing involves a complex interaction between a wide range of different processes, it places extremely high demands on the limited capacity of working memory. In order to avoid cognitive overload, writers have to develop effective strategies for managing the writing process (Flower & Hayes 1980b). In this paper, I will first outline the classical cognitive models of writing that embody these themes. I will then argue that more recent research has involved a shift in perspective – to a view of writing as text production – and sketch a dual process model of writing designed to
In the course of this review I will indicate some implications of the different models for research on writing in L2. My aim here is not to provide a detailed review of L2 research informed by cognitive models of writing, but rather to indicate how the different assumptions of these models might influence the questions and goals of research on L2 writing.

**2. The thinking behind the text**

Early research on writing was inspired by psychological research on problem solving. This provided a conceptual language for categorising the mental processes involved, a set of methods (verbal protocol analysis in particular) for examining these processes, and a body of empirical findings from research on problem solving in general which could be applied to understanding writing. It led to the development of a general model of the processes involved in writing (Hayes & Flower 1980), and to a theory of writing expertise (Hayes & Flower 1986). Hayes & Flower’s model distinguished between three basic processes: planning, which included generating ideas, organisation and goal setting as components; translating plans into text; and reviewing, which included reading and editing as components. These processes operated upon two kinds of information: a representation of the task environment, which consisted of the writing assignment and the text produced so far; and knowledge stored in long-term memory, which consisted of such things as topic knowledge, a model of the audience, the writing plan, rules for grammar production and knowledge of text standards. (Note, incidentally, that “translating” in this context refers to the process of converting conceptual content into a linguistic form, rather than to the process of translating from one language to another).

An important feature of the model, which distinguished it from a traditional product-based view of writing as a linear process of plan-write-edit, was the recursive nature of the process. Planning, translating and revising can, in principle, occur at any moment during writing – they refer to cognitive processes rather than stages in the writing process. The coordination of these processes was the responsibility of a monitor. The monitor in Hayes & Flower’s model therefore played a vital role in controlling the writing process - deciding when enough content had been generated,
when revision was necessary, and so forth. Individual differences in the way these basic processes were combined were attributed to different configurations of production rules representing the knowledge of the writing process stored in long-term memory.

Perhaps the most important consequence of this research was that it enabled a characterization of differences between expert and novice writers (Hayes & Flower 1986). Thus, Flower & Hayes (1980a) argued that experts construct a more elaborate representation of their goals, and continue to develop and modify this representation throughout the course of writing. In particular, they develop explicit rhetorical goals for the text as a whole, and use these to guide retrieval of content, whereas novices rely on more concrete content goals, and tend to generate content in response to the topic alone. In consequence, experts develop more elaborate plans, and continue to develop and modify these throughout the course of writing. In addition, the more elaborate conceptual representation of goals for the text enables experts to revise more extensively, evaluating their text in terms of its underlying function with respect to their goals, rather than simply considering whether the text is appropriately expressed (Hayes et al. 1987). Consequently, experts modify content more during both writing and revision.

Bereiter & Scardamalia (1987) summed up these differences between experts and novices as a contrast between a knowledge-telling model of writing and a knowledge-transforming model of writing. According to this model, the development of ideas during writing depends on the extent to which the retrieval of content is strategically controlled in order to satisfy rhetorical goals. Novice writers are assumed to employ a knowledge-telling strategy in which text production is guided by the direct retrieval of content from long-term memory and is organised solely by the associative relationships between content as it is stored in long-term memory. By contrast, more expert writers employ a knowledge-transforming strategy, which involves elaborating a representation of the rhetorical or communicative problem to be solved and using the goals derived from this representation to guide the generation and evaluation of content during writing. In consequence, more expert writers show much more evidence of reflective thought during writing: they develop more elaborate plans before writing, modify and elaborate these more radically during writing, and revise their initial drafts of texts more extensively. The end result is that more expert writers’ texts are tailored to the needs of the reader, and that in adapting their thought to their communicative goals, such writers also develop their understanding of what they are writing about.

© gfl-journal, No. 2-3/2009
Bereiter & Scardamalia (1987) formalised these differences in their knowledge transforming model of writing (see figure 1), stressing that this should not be seen simply as an evolution of the knowledge telling model but that it involved a radical change in the way that the writing task is defined by the writer and in the way that it is carried out. Thus, although it retains the knowledge telling model as a characterisation of the process whereby content is retrieved from memory, this is embedded within a dialectic between content and rhetorical problem spaces. This is intended to capture two features of the writing process. First, it reflects the fact that ideas are represented, not just as a reflection of the writer’s knowledge (content space), but also in terms of their rhetorical function within the text (rhetorical space). Second, writing is not simply a matter of adapting content to the rhetorical context, but is an emergent process in which content is formulated as the text develops. Thus, not only is content retrieved in response to a more elaborated representation of the assignment as a rhetorical problem, it is also formulated in the context of, and as a contribution to, the series of rhetorical acts gradually emerging in the text.

Figure 1: Bereiter & Scardamalia’s (1987) knowledge transforming model of writing.
The knowledge transforming model shares the general emphasis of classical cognitive models on the higher-level reflective thinking involved in writing. At first sight, the clear separation between thinking processes and text production processes made by these models might be taken to imply that the main focus of research on L2 writing should be on text production processes. One might assume that the goal-directed thought involved in effective writing is common to both L1 and L2 contexts, and that the essential difference between the two is in how the output of these central processes is formulated in language. However, a key feature of the knowledge-transforming model in particular is that it emphasizes the origin of the writer’s goals in their discourse knowledge. To the extent that L2 involves not just using a different language but also adopting different discourse conventions it may also involve learning different ways of thinking. A skilled L2 writer may find it difficult to adapt their writing process to an unfamiliar genre even when, and perhaps because, they are skilled and fluent writers in an L1 genre.

3. Cognitive overload

Aside from this difference in the goals towards which writing is directed, the other main factor emphasized in early models of writing was cognitive overload, arising from the fact that a complex set of processes has to be carried out in a limited capacity working memory. In particular, the demands of translating ideas into well formed text may consume resources required for higher level planning. Although this is particularly true for children, for whom even the basic mechanics of forming letters may be resource consuming, it is a pervasive problem stemming from the nature of the process itself.

A series of studies by Bourdin & Fayol comparing written and spoken recall with varying age groups suggest that low level processes involved in spelling and handwriting can also impair retrieval. In simple word-recall tasks, Bourdin & Fayol (1994) found that both second and fourth grade children recalled substantially fewer items when their responses were written (i.e. using relatively less practised handwriting and spelling skills) than when their responses were spoken (i.e. they could rely on more automatic speech production skills). There was no corresponding difference for adults. Similar results were found for a more complex sentence production task (Bourdin & Fayol 1996). However, when the composition task was substantially more complex
(Bourdin & Fayol 2002), they found that even adults perform worse in writing compared to speaking. This suggests that even when spelling and handwriting are very well practised, they can still have a residual effect on memory retrieval if resources are overloaded by other cognitively demanding processes.

The main implication of this general line of research is that it is important for other components of the writing process to be carried out as automatically as possible. Being able to write or type fluently and having well-developed language skills should reduce cognitive overload and facilitate more fluent retrieval of content from long term memory.

In addition, strategies for managing the writing process which help reduce cognitive load should also enable more effective planning. The most thorough investigation of the effectiveness of different drafting strategies was carried out by Kellogg in a series of experiments (Kellogg 1988, 1990; see Kellogg 1994 for a review). Kellogg (1988) compared the effectiveness of an outline strategy, in which writers generate and organize their ideas prior to writing before focusing their attention on translation and revision, with a rough-drafting strategy, which involves translating text without worrying about how well expressed it is, leaving monitoring of expression to revision of the draft after writing. There were two main findings. First, the strategies led to a redistribution of processing during writing (as measured by directed retrospection). In the outline conditions, writers planned less during text production, presumably because this had largely been completed prior to writing. In the rough draft conditions, revision was reduced during the initial draft and postponed until later. Second, although outlining was associated with higher quality final drafts, rough drafting showed no effect, despite the fact that revision had been postponed until after the initial draft. A later study by Kellogg (1990) suggested that the construction of a hierarchically organized outline prior to writing is associated with a higher quality final product than is the construction of an ordered list of ideas, and that this in turn is associated with higher quality final text than a simple clustering strategy. Kellogg’s (1994) general conclusion is that the effectiveness of the outlining strategy is a consequence of the fact that it enables writers to organize their ideas better prior to writing, as well as that it then enables them to devote more resources to formulating these ideas effectively in text.

The most obvious implication of these models for L2 research is that L2 language skills should have a strong impact on the writing process. Thus, L2 language proficiency would be expected to
affect not just how well-formed the written product is from a linguistic point of view, but also the writer’s capacity to engage in the higher level problem-solving activities characteristic of expert writing. Thus, even when L2 production is linguistically accurate, to the extent that L2 language production in L2 remains more effortful than in L1 one might expect writers to be less able to engage in goal directed creation of content and the quality of the text to suffer accordingly. It would be interesting to test, for example, whether writers in L2 showed similar decreases in their ability to retrieve content compared to retrieval in L1 as young writers do in retrieving content when writing compared to speaking (as in Fayol and colleagues’ research).

Furthermore, if the effort involved in L2 language processes does impair the ability of writers to engage in higher level planning processes then one would expect corresponding improvements in the quality of text produced under outlining conditions compared to single draft conditions, and that this would be a consequence of a reduction in the need to generate content at the same time as producing text (as in Kellogg’s experiments on effects of outlining in L1). This might seem a trivial replication of this research. However, in a recently completed experiment (Baaijen et al. 2008), in which we compared a group of students with dyslexia with a group of non-dyslexic students writing outlined planned texts, we found that outlining had very different effects for the two groups. For non-dyslexics, there was a strong negative correlation between the amount of content generated during text production and the quality of the text, as one would expect if the benefit of outlining is that it enables writers to separate content generation from full text production. By contrast, dyslexic writers wrote better texts the more they generated content at the same time as formulating the text. This could be because dyslexic writers’ difficulties with formulating text mean that, even when they are allowed to outline before they write, they still need to reconstruct content during text production. Alternatively, it could be because, for dyslexics, formulation in language is improved when it is carried out close to the point at which content is generated. Whatever the reason for the difference, the point for present purposes is that even the most robust findings of research on “normal” populations do not necessarily generalize to other populations. This is not to say, of course, that writing in L2 is like writing with dyslexia. But it does demonstrate the need to test even the most obvious implications of models of L1 writing in L2 contexts, and the need not to take them for granted.
4. From thinking to text production

Early cognitive models of writing, then, focused on the goal-directed nature of the thinking behind the text, and treated the translation of thought into text as a relatively passive component of the process, of interest primarily because a lack of fluency in translation was assumed to interfere with writers’ ability to engage in higher-level thinking. More recent research has begun to redress this balance and has paid much more attention to the processes involved in translation, and, in some cases has claimed a much more active role for it in the generation of content.

This is reflected in Hayes’ (1996) revision of the Hayes and Flower model, which makes much less clear cut distinctions between the different components of the writing process. Thus, planning has become one component of a more general ‘reflection’ module; translation has been renamed as text production, reflecting, perhaps, a less passive view of its role in content generation; and revision is treated, not as a separate process in its own right, but as a combination of the more basic processes of text interpretation, reflection and text production. In addition, working memory is explicitly incorporated into the model, and is characterised in terms of Baddeley’s multicomponent model of working memory rather than as general resource for which different components of the writing process compete.

In the same volume as Hayes’ revised model, Kellogg (1996) presented a model of working memory in writing, subsequently elaborated in more detail by Kellogg (2001). In Baddeley’s model (Baddeley 1986), working memory has three main components. The central executive is responsible for retrieval from long-term memory, control of attention, supervision of the system as a whole, and for coordinating the activity of the other two subsidiary systems. This central component is supported by, and controls the operation of, two ‘slave’ systems: the phonological loop, which stores and maintains verbal material in active memory, and the visuospatial sketchpad (VSSP), which stores and maintains visual and spatial material in active memory. According to Kellogg (see table 1), the planning component requires both the VSSP and the central executive but, since it is concerned with prelinguistic ideas, not the verbal component of working memory. The translation component requires the central executive to plan sentences and the phonological loop to store and maintain verbal material while sentences are being constructed. Transcribing language, which involves programming and executing motor routines, requires central executive resources, though this may be a minimal demand for practiced writers,
and such resources have minimal involvement in the executing as opposed to programming component of transcription. Reading previously produced text requires the central executive and the phonological loop, editing requires the VSSP because it involves maintaining an image of where text is on the page.

<table>
<thead>
<tr>
<th>Writing process</th>
<th>Component of working memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td>✓</td>
</tr>
<tr>
<td>Translating</td>
<td>✓</td>
</tr>
<tr>
<td>Programming</td>
<td>✓</td>
</tr>
<tr>
<td>Executing</td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>✓</td>
</tr>
<tr>
<td>Editing</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Kellogg’s (2001) model of working memory in writing.

The basic features of this model have been supported by a range of empirical research (see Kellogg 2001 and Torrance & Galbraith 2006, for reviews). In particular, two studies have suggested that the spatial component of working memory plays a crucial role in the development of new ideas (knowledge transforming) during outlining (Galbraith et al. 2005; Galbraith et al. 2009).

In a series of studies, Hayes (Chenoweth & Hayes 2001, 2003; Hayes 2009) has developed a more detailed model of the processes involved in text production and made comparisons of writers writing in L1 and L2. The model (see figure 2) consists of four components. The proposer is responsible for creating conceptual content – an idea package – which is sent to the translator. (For more extended texts, the proposer may involve goal-setting and other planning functions). The translator produces a language string which is then evaluated by the evaluator/reviser. If the string is acceptable it is passed to the transcriber to be turned into text. If the string is not acceptable, then the reviser can call on the other processes to produce a revised version of the
language or idea package, and this can in principle operate over a number of cycles before text is output. The reviser is also able to interrupt all other processes at any time.

![Diagram of text production model]

Figure 2. Chenoweth & Hayes’ (2003: 13) model of text production.

The model is designed to capture the fact that written language is typically produced in bursts of sentence parts rather than in complete sentences (Kaufer et al. 1986). These bursts often, but not necessarily always, consist of grammatical units. In particular, Chenoweth & Hayes contrast \(P\)-bursts, which end in a pause followed by further language production, with \(R\)-bursts, which are followed by revision of the language already produced. Their key claim is that the length (in words) of a \(P\)-burst depends on the capacity of the translator and this in turn depends on the writer’s linguistic resources. Essentially, the length of a \(P\)-burst depends on how much language the writer is capable of producing before the capacity limits of the translator are reached. For present purposes, their most important finding is that the length of \(P\)-bursts is reduced when writers write in L2 compared to L1 and that less linguistically proficient L2 writers produce shorter bursts than more proficient L2 writers. L2 writers also produce a higher percentage of \(R\)-bursts, i.e., they revise a higher percentage of the sentence parts they produce.

A key question this raises for L2 writing is what impact this has on a writer’s ability to formulate ideas in text. In earlier models of writing, the emphasis has been on the potential disruptive effect
of translation on global planning and reflection. In principle, this might be overcome by strategies like outlining. However, once one moves closer to the point of utterance, this is much harder to manage strategically. Ideas are often fleetingly generated at the point of text production and have to be maintained in working memory until the complete sentence has been transcribed. Clearly, how long it takes to complete the sentence, and the size of the parts that sentences are produced in, could have an impact on the ability of the writer to maintain the idea package they want to express in working memory. This could impact on the complexity of ideas that the writer is able to express and perhaps also on the local coherence of the text.

5. Writing as a knowledge-constituting process

According to the interpretation of Chenoweth & Hayes’ model that we have just considered, text production may affect not just the extent to which the writers are able to engage in higher level planning, but also the writer’s ability to capture fleeting thoughts as they occur, locally, in the course of text production. In a recently proposed dual-process model of writing, Galbraith (1999, 2009a, 2009b) goes further than this, and claims that spontaneous text production is an active knowledge-constituting process in its own right.

In a series of experiments investigating the conditions under which writers develop new ideas through writing, Galbraith and his colleagues (Galbraith 1992, 1999; Galbraith et al. 2006) have suggested that, although writers do develop their ideas more when they plan in note-form than when they try to produce full text at the same time as planning, as the knowledge-transforming model would predict, they also produce new ideas when they write spontaneous drafts of full text, and these ideas are associated with the development of the writer’s personal understanding of the topic. This has led to the development of a dual process-model in which effective writing is assumed to be the joint product of two conflicting processes. The first – knowledge retrieval – process involves retrieving already-formed “ideas” from an explicit store of knowledge in long term memory, and either translating these directly into text (what Bereiter & Scardamalia would characterise as “knowledge telling”) or the goal-directed evaluation and manipulation of ideas prior to translating them into text (what Bereiter & Scardamalia would characterise as “knowledge transforming”). By itself, however, this can only lead to the reorganization of
existing knowledge or to the selection of different items of existing knowledge which are more appropriate for the rhetorical context. In order to create new content, the writer has to engage in a different – *knowledge constituting* – process, which involves the synthesis of content guided by the connections between subsymbolic units stored in an implicit semantic memory system. Although this process can be prompted by higher level problem solving, the content produced by it is the product of the implicit organisation of content in semantic memory, rather than the explicit manipulation of content in working memory. Furthermore, because the units involved are subsymbolic – i.e. they are components of meaning and do not correspond to external referents – the meaning of what the writer wants to say only becomes apparent after the content has been formulated. The result is new content that can be added to the store of existing knowledge in explicit memory.

This model makes two claims about the knowledge-constituting process. The first is that, during text production, ideas are synthesized by constraint satisfaction within semantic memory, rather than being retrieved from episodic memory. In this respect, the model can be seen as a proposal about how the *proposer* component of Chenoweth & Hayes’ (2003) model produces the “idea package” that serves as the input to the *translator* component. The main consequence of this way of conceiving of the *proposer* is to emphasize the transient nature of ideas during text production: they are not fixed ideas retrieved from long-term memory but are temporary patterns of activation across the set of units constituting the writer’s semantic memory.

The second, more radical, claim is that a sequence of utterances need not necessarily be the product of explicit planning in between syntheses of content. Galbraith (1999) suggests that when inhibitory feedback from a previous utterance is input to semantic memory it reduces the activation of units corresponding to the preceding utterance so that, without any change in the writer’s goals, subsequent syntheses will correspond to the “remainder” of the content implicit in semantic memory. This allows thought to be “self-moving”, with each successive utterance causing subsequent utterances. There are two key features to this. The first is that, because the writer does not have direct access to the constraints within semantic memory that guide the synthesis of content, they only become aware of the content of any given utterance at the moment it is created. Second, because any given utterance is only a partial representation of the content of semantic memory, in order to capture the content implicit in semantic memory, the writer has to allow the process to unfold without interruption by explicit planning. Their understanding is
constituted by the interaction between successive utterances and the implicit content of semantic memory, and in order to articulate it they have to allow the process of text production to unfold without interruption.

This characterisation of text production as a knowledge-constituting process has an important implication for writing in L2, arising from the fact that language is produced in bursts and that the size of these bursts appears to be reduced in L2 (Chenoweth & Hayes 2003). If, as the knowledge-constituting model claims, these bursts play a constitutive role in the development of the writer’s understanding, then the reduced size of the bursts in L2 should alter, and perhaps, reduce the extent to which writing in L2 leads to such developments. This could be tested by replicating the measures used in Galbraith’s experiments, and comparing the extent to which writers develop their understanding in L1 and L2.

A second important implication arises from the conflicting nature of the two sources of content organization assumed by the dual-process model, both of which are required for effective writing. The knowledge-retrieval process organises content in terms of the relationships between pre-existing ideas in explicit memory and the writer’s rhetorical goals, whereas the knowledge-constituting process is guided by the implicit organisation of the writer’s semantic memory. Galbraith (2009a) suggests that this is not simply a cognitive conflict. It is intimately related to the writer’s conception of self. The priority that the writer gives to the two processes depends on the extent to which they are motivated to present a coherent self-image to the reader (through goal directed planning) or to actualize the potential self latent in their implicit disposition towards the topic (through spontaneous text production). Writing in L2 may affect the balance between these two processes in a number of ways. On the one hand, to the extent that it is a more self-conscious process than writing in L1, it may lead the writer to prioritise explicit planning processes more than they would in L1. This is not to say that these would be carried out more extensively, but rather they may shift their attention to satisfying more formal constraints on the text at the expense of being concerned with the extent to which the text captures and articulates their personal understanding. On the other hand, to the extent that the writer finds it harder to articulate their personal understanding in L2, their motivation to write may be reduced. If one of the factors that motivate writers is the sense that they are developing their understanding, then any reduction in their capacity to do this may reduce their motivation to write. There is a dialectical relationship between cognitive and social/motivational processes: the processes
employed by writers affect their motivation, and their motivations influence the processes they employ. Cognitive processes in L2 writing cannot be studied separately from the social and motivational contexts in which they occur.

6. Conclusion

The main point that emerges from this brief review of research on writing processes is that learning to write in a different language is not just a matter of developing more fluent linguistic skills. It is not a matter of taking thoughts in one language and trying to translate them into the words of another language. Writing is thinking, and it is the effects of L2 on the writer’s thoughts as they try to write that need to be researched. This review suggest a number of questions. What different genre conventions are there in L2 contexts and how does the writer’s understanding of these impact on their ability to write in a goal-directed and purposeful manner? How does fluency in L2 impact on the writer’s ability to carry out higher level thinking processes, and what sorts of strategies might enable them to do this better? Are these necessarily the same as the kinds of strategies that have been found to be effective in L1 contexts? How do differences in both linguistic fluency and linguistic structure affect the writer’s ability to constitute their thought in writing? Research in L1 may have helped us to develop these questions, but only research on L2 writing itself can help us answer them.

References


© gfl-journal, No. 2-3/2009


Biographical data

David Galbraith is Senior Lecturer in Psychology and Director of the Centre for Educational Psychology Research at Staffordshire University in the UK. He researches into cognitive processes in writing (particularly the effects of writing on cognition) and into the use of writing in educational and therapeutic contexts. He has been coordinator of the European Association for Learning and Instruction's Special Interest Group on Writing and is currently a committee member of the European COST Action for Learning to Write Effectively.