Units of production in writing: evidence of topic 'framing' in on-line writing research

Kristyan Spelman Miller

School of Linguistics & Applied Language Studies, The University of Reading

Abstract. This paper is concerned with the definition of pause location in the analysis of on-line text production. A basic concern in psycholinguistic and applied linguistic research into language production is the selection of a suitable and valid unit on which to base the measurement of productivity and fluency. In existing research on spoken language, units have tended to be defined according to syntactic, semantic and phonological criteria, but do not embrace discourse-related notions such as topic which could be important in the analysis of the production process. In my recent work on written text production using keystroke logging, I have attempted to extend and develop existing notions of units of production in the direction of the potential discourse roles of units of language. In this paper I outline the set of so-called 'framing devices' proposed to reflect the thematic function of certain units, and illustrate these categories through data from a study of L1 and L2 writers of English within an academic context.

1. Introduction

The aim of this paper is to present a descriptive framework for the identification of pause location, developed as part of a larger scale keystroke study of L1 and L2 writing processes. The central importance of the choice of a unit of measurement of language output has exercised researchers concerned with the description and quantification of language production for many decades. It also underlies a number of more recent review articles such as Crookes (1990) and Foster et al (2000) which outline a range of alternatives available to the researcher, including the utterance/sentence, c-unit, tone-unit, T-unit, idea unit, and AS-unit. Clearly, it is on the basis of some unit of production that statements of quantity of production, frequency of certain features, and measurements of accuracy, complexity and fluency may be expressed, and that attempts may be made to compare output in a variety of conditions. Decisions concerning the way language output is characterised, based on such different definitional criteria as syntactic, semantic or phonological form, are therefore far from inconsequential and need to be given careful consideration and critical evaluation.

Whereas much of the traditional research interest has been on the description of spoken language output, it is no less important in the case of the analysis of written text production - the so-called 'stepchild of psycholinguistics' (Bonin & Fayol 1996:145) - to reflect on the range of options available for the measurement and description of language output. In recent decades the increased interest in researching the on-line processes of writing puts into particular relief the need for a well-defined means for describing the characteristics of the language as it is produced. It is this concern with the textual/cognitive dimensions of writing which underpins this current paper.

The need to consider and refine existing approaches to the description of written language production arose from my immediate needs in handling data from a keystroke study of L1 and L2 writers. In brief, this approach involves the unobtrusive recording of writers as they compose on a computer (see Spelman Miller 1999, 2000a, 2000b for fuller descriptions of the approach). Throughout the writing event, resident software records all operations made in real-time, and stores information electronically in logfiles for subsequent analysis. The output of the logfiles is a highly detailed record of writing activity, giving access to information about the temporal features of writing on-line and the sequence of processes such as planning, revising and formulating text.

The study we report here is concerned with the fluency and productivity of L1 and mixed L2 writers working in English, producing two academic writing tasks differing in terms of rhetorical demand (descriptive versus evaluative essay). This two-by-two design allows us to investigate the potential effect of task and of language group on aspects of writing production. For the purposes of this paper, we focus on such temporal features as:

- **pause duration**: instances of keyboard inactivity (greater than two seconds) measured in seconds and tenths of seconds;
- **pause frequency**: instances of inactivity identified between keyboard activity such as character presses or operations (such as deleting, scrolling); and
- **productivity (length of text span)**: the length of text produced between two pauses.

A fundamental step in the analysis of these variables is the characterisation of the text produced in terms of strings or units of language output. As text is produced in real time, what spans of text

emerge? Do they conform to (written) grammatical units such as word, phrase, clause, sentence, paragraph level constituents, or are other characterisations (for example, discourse related units) feasible and interesting? In the sections which follow we outline the framework devised for the description of this language output, drawing attention to ways in which the categories diverge from or move beyond characterisations used in the general psycholinguistics and writing process literature. In particular, in addition to the grammatical categorisation favoured in much of the existing literature, we introduce a discourse-oriented unit, referred to as a framing device, by which we aim to explore the units of language produced in terms of their potential function to introduce, maintain, or shift topic. We then illustrate the application of these framing device units to the data elicited in the project, and comment on the insights revealed from such an analysis.

2. Traditional approaches to the definition of pause location

The work in mainstream psycholinguistic research over previous decades on the identification of pausing in speech (Henderson et al 1966, Goldman-Eisler 1968, Butterworth 1980, Beattie 1983, and so on), has established a long tradition of concern with the correspondence between location of non-fluency (associated with planning) and the structural composition (essentially grammatical) of the spoken message. Such work has focussed in particular on the occurrence of greater non-fluency at clause and sentence boundaries (Goldman-Eisler, 1972, for example). Other work, too, has drawn attention to other constituents in the string of speech which attract pauses, for example, following an utterance-initial connective, and before an adverbial constituent and before lexical (content) items (Garman 1990).

Whereas this work on spoken language production clearly puts an emphasis on the grammatical / semantic characterisation of the language produced, other approaches, such as that taken by Grosjean & Deschamps (1975), Grosjean, Grosjean & Lane (1979) and Gee & Grosjean (1983), reveal other, prosodic influences on the clustering of constituents (or 'performance structures') as text is read aloud. The correspondence between units of production and prosodic structure is also the focus of attention in a range of other discussions, such as Halliday (1967), Boomer (1965), Chafe (1980), Brown and Yule (1983). It is clear from this, then, that in the case of spoken language production researchers draw on a range

of criteria, be they syntactic, semantic or phonological, to identify, segment and analyse their data.

In the study of real-time writing, researchers have tended to favour the grammatical characterisation of pause location and text span produced. Matsuhashi (1981, 1982, 1987), Warren (1996), Janssen et al (1996), Schilperoord (1996), Stromqvist & Ahlsen (1998), and others, have made use in a variety of ways of such grammatical units as the sentence, clause, T-unit, phrase, word and character. At this stage in the development of on-line writing studies, however, it might be timely to review and evaluate the means by which the textual data are described and analysed. In particular we highlight three main concerns which inform the direction of our own work reported in this paper:

- There is a need to refine the categorisation of items at word level, which the general label of word fails to differentiate (for example, items of different classes: determiner, adjective, noun, adverb, disjunct, conjunction, and so on);
- The dynamic, evolving nature of the on-line data needs to be accounted for in the description of the units produced. In other words, location needs to be seen in terms of potential as well as actual (produced) units of language.
- Grammatical characterisations do not reflect features of the discourse structure of the text. The innovative work of Sanders et al (1996) and Schilperoord & Sanders (1999) is different in this respect. They consider the association between text production and topic development (continuity and discontinuity), by relating pausological data to the hierarchical arrangements of segments connected by clause relations such as claim-argument, problem-solution and sequence. Such a discourse-oriented approach inspires the development of our own framework for analysis.

It is in response to these three concerns that we propose a framework to be presented below. We begin with the main features and categories of our scheme in general before focussing on the discourse-oriented approach taken in the characterisation of units of production.

3. General features of the grammatical framework

As an initial starting-point for the characterisation of the units of written text production, we choose to categorise pauses according to their

grammatical locations within the stretch of text. Based on a Hallidayan concept of structure (Thompson, 1996; Halliday 1985, 1994) which identifies functionally distinct elements within the stretch of text, we distinguish the following basic units for the definition of pause location: sentence, clause, intermediate constituent (phrase/group), word, and character (morpheme). The identification of these units is based on content criteria as defined later (see Figure 2 below).

In order to accommodate the need to describe the data as arising from a dynamic process, that is, with the text string potentially subject to alterations of form as the text emerges, we identify pause location in terms of the <u>preceding</u> structural element or elements. In other words, categories of structure used to describe location reflect the status of the unit at a particular point in the construction of the text, in terms of potential rather than definitive or static structural units. To give an example, a pause immediately following the noun phrase *research* should be characterised as an intermediate constituent, regardless of subsequent modifications to the form of the item (which may alter *research* to *researchers* or *researched*, for example)[1]. Locations are therefore referred to as *potential completion points* to reflect the idea that the development of the text is not predetermined but open to adjustment in a number of different ways as the writer reworks the text.

The following categories are therefore identified for the analysis of pause location:

- sentence completion point (SCP)
- clause completion point (CCP)
- intermediate constituent (or phrase/group) completion point (ICP)
- word completion point (WCP)
- character completion point (XCP)

The categorisation is designed as a working model to capture the main locations of interest in our data, and cannot claim to address all possible occurrences. The procedure for analysing is summarised in a flow-chart (see Figure 1) as a series of decisions with respect to potential completion, and the specific grammatical composition of these units further defined in Figure 2. Such a characterisation helps us to differentiate more sensitively functionally distinct units.

---

[1] It may be the case of course that the occurrence of a space before the pause will help to confirm its status at a word (or phrase group) versus character-level location.
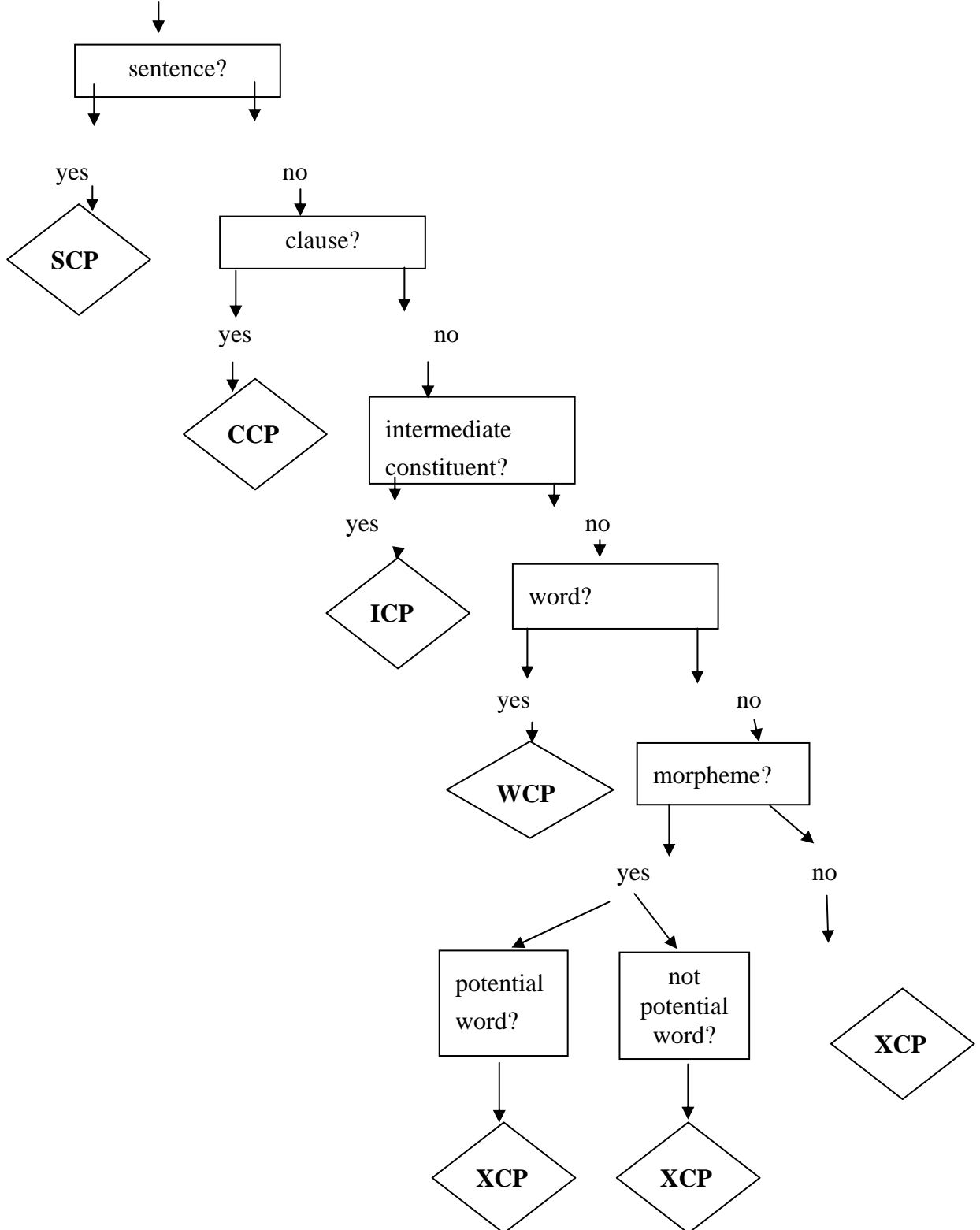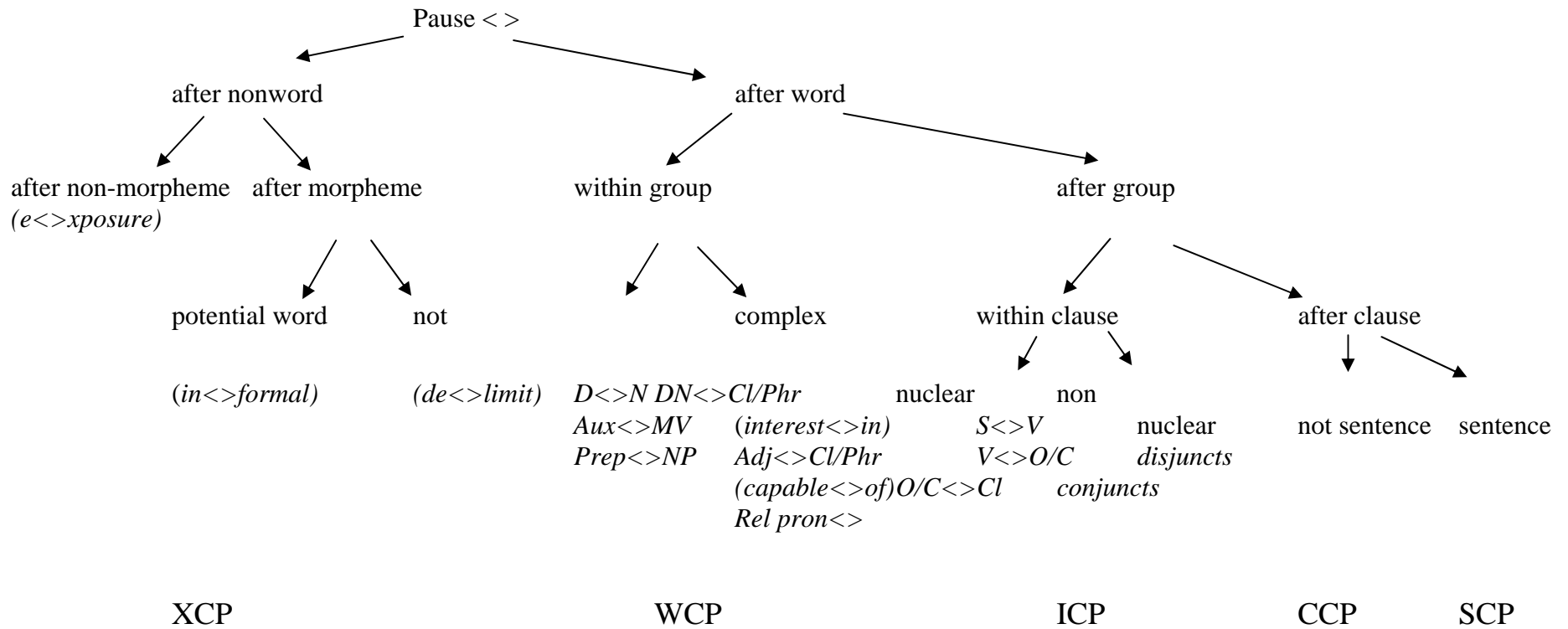
Pause < > at completion point

Fig. 1 Identification Procedure

Pause < >

after nonword   after word

after non-morpheme   after morpheme   within group   after group
(e<>xposure)

potential word   not   complex   within clause   after clause

(in<>formal)   *(de<>limit)*   *D<>N DN<>Cl/Phr*   nuclear   non
                              Aux<>MV   *(interest<>in)*   *S<>V*   nuclear   not sentence   sentence
                              Prep<>NP   *Adj<>Cl/Phr*   *V<>O/C*   *disjuncts*
                                     (capable<>of)O/C<>Cl   *conjuncts*
                                     Rel pron<>

XCP   WCP   ICP   CCP   SCP

Key: D = determiner; N = noun; Aux = auxiliary; MV = main verb; S = subject Prep = preposition; NP = noun phrase; Cl/Phr = clause/phrase; O/C = object/complement; rel pron = relative pronoun; Adj = adjective; (examples given in brackets).

Figure 2: Identification of pause locations (main category contents)

4. The discoursal framework: 'framing'

One of our major goals, however, has been to move beyond the solely grammatical characterisation of units of production, to reflect certain discoursal features of the language produced. In order to do so, we reconsider the units presented in the grammatical analysis from the point of view of a number of functions related to the introduction and development of topic in the discourse. In particular, elements which occur in the intermediate constituent category (nominal groups in subject or adjunct position, and also conjuncts and disjuncts) and certain clause-level elements may be reinterpreted as performing particular functions with respect to framing or setting up the rest of the message. These units are identified as instances of **framing device**. The choice of this term, framing device, is influenced by Witte & Cherry (1986), following Bracewell et al (1982), who refer to processes by which the writer makes choices about topicalisation in clauses and the establishment of topical relations across clauses (Witte & Cherry 1986:127).

The notion of topic is too complex to discuss in detail here. Suffice it to say that decisions have had to be made about the definition and scope of our treatment of this term with respect to our data. In brief, we are concerned with topic at a local (sentence) rather than discourse level, since this is the level with which much of the psycholinguistic (production processes) literature in general and pausological research in particular has been concerned.

A further justification for this local-level approach to the description of topic comes from the on-line nature of our data. Since our analysis is concerned with describing text at the point of production, we are not in a position to know the consequences or effect in global, discourse terms of the selection of certain units in the text string. Indeed, an element may not survive phases of revision to appear in the final text product, or its function may not be clear at the point at which it is produced. Equally, the contribution of an item in the text string to the development of the discourse topic may not become clear until the whole text is produced. Since our study aims to describe text units as they emerge at the point of production (and without knowledge of their ultimate effect on the overall text structure), we set the scope of our consideration of topic, therefore, at the level of sentence/clause.

In order to investigate topic from this sentential viewpoint, we propose the notion of a set of 'framing devices', which are associated with topic-related functions (see Goutsos (1997) for similar approaches to written product). Put simply, a framing device is an element or structure

(single word, phrase or clause) which serves to establish the starting point of the message at the clause/sentence level. This may be in one of a number of ways: in constituting the topic itself, or in preparing the scene for the introduction or change of the topic, for example through the use of a discourse marker such as *however*, empty theme such as *it is (stated) that*, or initial clause structure such as *By doing this*.

---

1.   Subject theme
<u>This hypothesis</u> might be attributed to the claim of lateralisation
<u>This</u> would be further discussed

2.   Adjunct theme/complement theme
<u>Around puberty</u>, human beings will face with lateralisation of the brain
<u>Among individual factors</u>, those which are widely recognised by most scholars are: age aptitude, attitude, motivation and personality

3.   Non-experiential theme
<u>In addition</u>, Long (1990) provides evidence to suggest that acquisition of a native-like accent is not possible after the sixth year
<u>To start with</u>, in an attempt to present a theoretical view of motivation, Skehan put forward four hypotheses

4.   Empty theme (*it*, *what* and existential *there*)
<u>There are</u> debates on the methodology of experiments carried out
<u>What is needed in this area of research</u> is some longitudinal studies
<u>it is that</u> young learners acquire a language more easily
<u>it is beyond doubt that</u> everyone apart from some exceptions can learn a language

5.   Thematised structure (eg finite/nonfinite clauses)
<u>Because of a number of individual differences in the same age group</u>, there should be different processes of L2 learning and acquisition
<u>Since I was a child</u>, his big dream was I to become an English teacher just like him

---

Fig. 3 Categories of framing device

The taxonomy we present (see Figure 3) consists of five types of framing device which fulfil these topic-related functions. They are robust, low-inference categories which can be applied to the existing pause location categories (ICP and CCP) described above.

To conclude this presentation of our framework for pause location analysis, then, we point to the three main issues we seek to address. Firstly, we present a linguistically more sophisticated approach to the identification of certain items in the text string in an attempt to characterise more subtly differences between certain word classes. Secondly, we represent location in terms of potential completion points, as a more appropriate means of accommodating the emergent and often temporary status of text elements as they are produced on-line. Thirdly, the notion of location is extended to incorporate a discourse-sensitive category, the framing device, which allows us to interpret certain text units in terms of topic-related functions.

In the section which follows, we refer to the application of these categories to the study of writing processes and textual output. This will allow us to speculate on the type of insights such an analysis might bring to the investigation of writing behaviour.

5. 'Framing' applied to a study of L1 and L2 writing

As we have mentioned briefly above, the development of this framework arose in the context of a larger study of L1 and L2 writing processes, which is reported more fully elsewhere (Spelman Miller 1999, 2000a, 2000b). In brief, the study is concerned with the writing processes of a number of L1 and mixed nationality L2 writers (n=10, and n=11, respectively) producing academic essays under timed, simulated examination conditions. The two tasks given to the subjects varied in rhetorical demand (descriptive versus evaluative) but were in general terms based on the same topic area. The broad aim of the study was to compare performances of subjects on the two tasks in terms of the pause-related phenomena outlined above (pause duration, pause frequency and rate, productivity and rate of production). Subjects for the study were volunteers from a larger group of students participating in a University lecture course, and were matched in terms of age, previous knowledge of the topic area, and experience of composing using a computer. Participants' motivation to the task was considered high given the nature of the task as preparation for their final university assessment.

Quantitative analysis of the writers' performance in terms of the main dependent variables stated above needs to be cautiously interpreted given the small number of subjects involved in the study. However, it is possible to report clear overall findings with respect to the predicted differences between the L1 and L2 writers. The L2 writers produced longer pauses at all grammatical locations, and especially at word-internal and intermediate constituent locations, suggesting increased attention to lower-level processing concerns. Productivity and rate of production were also significantly lower in the case of the L2 writers. On all measures, however, the writers behaved in very similar ways under the two task conditions, suggesting a lack of perception of, or inability to respond to, the differing rhetorical demands of the two tasks.

Of particular interest to our present discussion is the application of the framing device unit to data in our study. As we have outlined above, this exploratory unit has been proposed as a means of capturing structures in the text string associated with the establishment, maintenance and development of the topic. The selection of exponents of this framing device category is far from definitive, but includes a number of key units such as subject theme, non-experiential theme (including disjuncts and conjunctions), and thematised structures, that is, fronted clausal and phrasal structures. Indeed, our initial observations reveal that these three are by far the most frequently occurring framing device categories, with subject theme being the most prominent category. In fact, much of the quantitative analysis of framing device occurrence (as reported below) is based on the subject theme category alone.

Our analyses so far seem to support the notion of the framing device as a potentially interesting means of interpreting the data string produced during composition. In general terms, we found an interesting coincidence of framing device and pausing: approximately one third of framing device locations coincided with the boundary of a unit of production (i.e. a pause). This suggests a potentially powerful role for these framing devices in the writer's management of the formulation process.

In more specific terms, the analysis of framing devices appears to open a window on the subtle processing preferences of writers when producing text. Investigation of pause duration at framing device locations reveals both general significant differences between the subject groups and individual differences between writers, which we summarise briefly in sections 5.1 and 5.2 below.

5.1 Framing and L1 / L2 writer groups

Although no evidence of statistically significant difference is found in overall framing device occurrence across the two language groups, when the most frequent category, subject theme, is considered, some interesting observations may be made. Firstly, although patterns of pausing at such locations vary according to individuals, the general impression is that the (full noun phrase) subject theme is a likely location for pausing.

Secondly, with respect to pause duration at the framing device location, through comparison of pause length at subject theme and non-subject theme (but equivalent ICP) locations, we are able to report that a significant effect of language group on pause duration at subject theme locations ($p=0.001$) but not with respect to non-subject theme locations. The nature of the inter-group differences with respect to pause duration at subject and non-subject theme locations is further clarified in Table 1 below.

	subject theme	non-subject theme
L1 mean	5.13	5.24
st dev	*0.44*	*0.34*
L2 mean	7.26	6.16
st dev	*0.42*	*0.33*

Table 1 Mean pause durations (in seconds) at subject theme
and non-subject theme ICP locations by language group

This information reveals that in the case of the L1 writers, means for subject theme and non-subject theme pauses are very similar (5.13 and 5.24 seconds respectively), but for the L2 writers, pauses are markedly longer at subject theme locations than at the non-subject theme ICP locations (7.26 and 6.16 seconds respectively). This points once again to inter-group differences with respect to pause behaviour. A possible interpretation of these findings is that the L2 writers appear to make use of the subject theme framing device location to produce longer pauses, whereas in the case of the L1 writers, in general, the subject theme location does not attract substantial pausing. Indeed, in the case of the L1 subjects mean durations of subject theme pauses are slightly lower than those of other (non-subject theme) ICP pauses.

Such findings ought to be carefully handled, however, given the degree of individual variation which we note in our data. We might therefore prefer a more tentative observation, that for certain individuals

within the L2 group the framing device slot appears to coincide with marked pausing. A finer grained exploration of individual writing episodes, as we propose below (section 5.2), would help to elaborate the different patterns of pausing of individual writers at this and other locations in our analysis.

5.2 Framing and individual writers

Case study analyses of three individual L2 and L1 writers allow us to pursue this line of enquiry further. From the detailed analysis of each case, in which we consider pause location at a number of key locations for the development of the discourse (namely, CCP/SCP and framing device locations), we are able to report some intriguing and complex differences between individuals. The subjects we investigate are a Japanese writer of English L2 (L2:A), a Greek writer of English L2 (L2:Ya) and an L1 English writer (L1:E).

The incidence of pausing at framing device locations on both tasks, D and E, (as a proportion of possible uptake) is reported in Table 2. We note that, in general, approximately one third to one half (27% to 48%) of all potential framing device slots coincide with pausing, which suggests that this location is for some writers highly susceptible to pausing, although this may vary according to individual preferences.

Subject	Task D	Task E
L2:A	0.45	0.41
L2:Ya	0.48	0.31
L1:E	0.27	0.37

Table 2 Proportion of instances of framing device categories associated with pausing

The uptake of pausing at subject theme location (see Table 3) reveals a similar picture. Between a quarter and a half of all such available slots coincide with pausing, although individual subjects seem to differ in their 'use' of these slots for pausing. L1:E makes limited use of the framing slot location, preferring to pause at inter-clause/sentence locations. L2:A, by contrast, uses the framing device location to a large extent on both tasks, and more so than her L2 counterpart, L2:Ya. Such observations suggest on the one hand consistent behaviour on the two tasks, but on the

other quite radical differences in the strategies shown by individual writers.

When mean durations of pausing at framing device locations are compared with those at CCP/SCP locations, further insights may be drawn into the features of individuals' writing strategies. Table 4 presents the mean durations of pauses at these locations for each individual on the two writing tasks. A final column presents the number of framing device pauses as a proportion of total pauses.

Subject	Task D	Task E
L2:A	0.49	0.51
L2:Ya	0.35	0.27
L1:E	0.25	0.31

Table 3 Proportion of instances of subject theme framing devices associated with pausing

This information suggests considerable variation between individuals in pause duration. L2:A appears to pause for the least amount of time at all these locations, and on both tasks. Particularly striking is the relatively low mean duration at the SCP location, which in the case of the other two subjects, is markedly higher than the figure for the CCP location. L2:Ya produces extremely high mean pause duration at SCP locations on Task D (with an extremely high standard deviation); for L1:E mean pause duration is more consistent across the two tasks (mean = 11.05 and 11.26 secs), but is still higher than at the CCP location.

When mean durations at framing device locations are added to the picture, similar differences emerge between individuals. In general, pauses at framing device locations account for a similar proportion of all pauses produced for all subjects (from 14% to 26%), although there are differences between individuals in mean pause lengths at these framing device locations. In the case of L2:A, pauses are shorter than for the other subjects. For L2:Ya, however, pauses at these slots tend to be longer, particularly so in the case of subject theme pauses. In other words, although the uptake of framing device slots by L2:Ya is less regular, when pauses do occur here they tend to be particularly lengthy. The L1 writer, L1:E, produces pauses at subject theme slots of similar duration to those of L2:A, but when durations are considered for all categories of framing device, pauses are on average considerably longer.

In general terms, then, we may summarise our findings as follows. For the two L2 writers, but not for the L1 writer, the framing device location appears to be a natural location for pausing, although the nature of the pause behaviour differs. In the case of one L2 writer (L2:A), subject themes are particularly susceptible to pausing, although pauses at these locations tend not to be lengthy. For the other L2 writer (L2:Ya), pausing is less frequent at these framing device locations but when it does occur pauses tend to be very long. Finally, in the case of the L1 writer (L1:E), extensive pausing occurs almost exclusively at clause and sentence locations and not at framing device locations.

Subject	CCP (secs)	SCP (secs)	All framing devices (secs)	Subject themes (secs)	Proportion of fds to all pauses
L2:A_D	4.42	5.84	4.99	4.12	0.26
L2:A_E	5.63	6.96	5.25	3.84	0.21
L2:Ya_D	6.53	23.12	6.66	5.92	0.26
L2:Ya_E	5.81	8.78	7.00	9.43	0.16
L1:E_D	6.16	11.05	6.11	4.16	0.14
L1:E_E	7.12	11.26	6.35	4.26	0.22

Table 4 Mean durations of pauses (in seconds) at key locations

The close analysis of individual pausing behaviour, which may be pursued at increasing levels of detail - for example, the analysis of the interaction between framing device occurrence and the length of text span produced, or preceding revision activity - potentially offers insights into differences in text production strategy. By demonstrating the application of our analytical categories to specific data, we are able to explore the complexities of individual processing styles, and, in so doing, open up for critical examination the categories we have established for describing and interpreting pause behaviour.

6. Conclusions

The observations we have been able to make on the data from our study, while being tentative and speculative, based on a small sample of writers,

allow us to glimpse the complexity of the writing process viewed from a cognitive-textual perspective. Moreover, in general terms, they highlight the need for the careful definition and application of units of analysis, and encourage the refinement and extension of the type of framework we have offered here for the description of units of written text production.

While the goal of this paper has been to exemplify a scheme for the development of categories of analysis in on-line writing studies, we are aware that we have not had space to explore the application of such a concept of location to the study of aspects of writing behaviour, such as productivity and revision. This we leave as a prospect for further work. For the future, too, we point to the need to develop the notion of framing device through the collection of larger data samples. We are conscious that we have focussed almost exclusively on the subject theme category, because of limitations of the amount of data obtained. Larger sample sizes would elicit more extensive coverage of other framing device categories.

Finally, the concept of the framing device itself could be developed and refined, for example, from the perspective of socio-contextual concerns with authorial stance, or writer-reader roles. We are conscious that the cognitive-textual dimension we have selected for our analysis is partial, and lacking connection to broader socio-contextual matters underlying writer choice. For the time being, however, we offer this exploratory approach to the discoursal interpretation of pause location and units of production for evaluation and improvement in subsequent research.

References

Beattie, G. W. (1983). *Talk: an analysis of speech and nonverbal behaviour in conversation*. Milton Keynes: Open University.

Bonin, P. & Fayol, M. (1996). Producing isolated words from pictures, from orally and visually presented words: an on-line study of naming and writing. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (eds.) *Current research in writing: theories, models and methodology*. Amsterdam: Amsterdam University Press. 145-158.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech* **8**. 148-158.

Bracewell, R., Frederiksen, C. & Frederiksen, J. (1982). Cognitive processes in composing and comprehending discourse. *Educational Psychologist* **17.** 146-164.

Brown, G. & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.

Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (ed.) *Language production: Vol. 1. Speech and talk.* London: Academic Press. 155-176.

Chafe, W. (1980). The deployment of consciousness in the production of a narrative. In W. Chafe (ed.) *The pear stories: cognitive, cultural and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

Crookes, G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied Linguistics* **11.** 183-199.

Foster, P., Tonkyn, A. & Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics* **21.** 354-375.

Garman, M. (1990). *Psycholinguistics.* Cambridge: Cambridge University Press.

Gee, J. & Grosjean, F. (1983). Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology* **15.** 411-458.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* London: Academic Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech* **15.** 103-113.

Goutsos, D. (1997). *Modeling discourse topic: sequential relations and strategies in expository text.* Norwood, NJ: Ablex.

Grosjean, F. & Deschamps A. (1975). Analyse contrastive des variables temporelles de l'anglais et du francais: vitesse de parole et variables composantes, phenomenes d'hesitation. *Phonetica* **31.** 144-184.

Grosjean, F, Grosjean, L. & Lane, H. (1979). The patterns of silence: performance structures in sentence production. *Cognitive Psychology* **11.** 58-81.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English, parts 1-3. *Journal of Linguistics* **3.** 37-87, 199-244, **4.** 179-215.

Halliday, M. A. K. (1985). *An Introduction to functional grammar.* London: Edward Arnold.

Halliday, M. A. K. (1994). The construction of knowledge and value in the grammar of scientific discourse, with reference to Charles Darwin's *The origin of species.* In M. Coulthard (ed.) *Advances in written text analysis.* London: Routledge. 136-156.

Henderson, A., Goldman-Eisler F. & Skarbek (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech* **9.** 207-216.

Janssen, D., van Waes, L. & van den Bergh H. (1996). Effects of thinking aloud on writing processes. In C. M. Levy & S. Ransdell (eds.) *The science of writing.* Mahwah: Lawrence Erlbaum Associates. 233-250.

Maclay, H. & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word* **15.** 19-44.

Matsuhashi, A. (1981). Pausing and planning: the tempo of written discourse production. *Research in the Teaching of English* **15.** 113-134.

Matsuhashi, A. (1982). Explorations in the real-time production of written discourse. In M. Nystrand (ed.) *What writers know: the language, process, and structure of written discourse.* New York: Academic Press. 269-290.

Matsuhashi, A. (1987). *Writing in real time: modelling production processes.* Norwood: Ablex.

Sanders, T., Janssen, D., van der Pool, E., Schilperoord J. & van Wijk C. (1996). Hierarchical text structure in writing products and writing processes. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (eds.) *Current research in writing: theories, models and methodology*. Amsterdam: Amsterdam University Press. 473-492.

Schilperoord, J. (1996). The distribution of pause time in written text production. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (eds.) *Current research in writing: theories, models and methodology*. Amsterdam: Amsterdam University Press. 21-35.

Schilperoord, J. & Sanders, T. (1999). How hierarchical text structure affects retrieval processes: implications of pause and text analysis. In M. Torrance & D. Galbraith (eds.) *Knowing what to write: conceptual processes in text production*. Amsterdam: Amsterdam University Press. 13-30.

Severinson Eklundh, K. & Kollberg P. (1996). A computer tool and framework for analyzing online revisions. In C. M. Levy & S. Ransdell (eds.) *The science of writing*. Mahwah: Lawrence Erlbaum Associates. 163-188.

Spelman Miller, K. (1997). Investigating writing on-line: an illustration of methodology and analysis. *Reading Working Papers in Linguistics* **3.** 323-340.

Spelman Miller, K. (1999). Investigating academic writing on-line; a study of temporal aspects of text production. In M. Pennington (ed.) *Writing in an electronic environment: research with language learners.* Houston, Texas: Athelstan.

Spelman Miller, K. (2000a). Academic writers on-line: investigating pausing in the production of text. *Language Teaching Research* **4.** 123-148.

Spelman Miller, K. (2000b). *Writing on-line: temporal features of first and second language written text production*. PhD thesis, University of Reading, UK.

Stromqvist, S. & Ahlsen E. (1998). *The process of writing – a progress report.* (1st draft). Department of Linguistics and SSKKII Center for Cognitive Science, University of Goteborg, Sweden.

Thompson, G. (1996). *Introducing functional grammar*. London: Arnold.

Warren, E. (1997). The significance of pauses in written discourse: a comparison of native speaker and learner writing. In A. Archibald & G. Jeffery (eds.) *Second language acquisition and writing: a multidisciplinary approach*. The University of Southampton. 152-177.

Witte, S. & Cherry R. (1986). Writing processes and written products in composition research. In C. Cooper & S. Greenbaum (eds.) *Studying writing: linguistic approaches*. Newbury Park, CA: Sage. 112-153.