# Functional Grammar Processing for English & Chinese Texts

Jonathan J Webster

Principal Lecturer
Department of Applied Linguistics
City Polytechnic of Hong Kong

#### Introduction

The Functional Grammar Processor (FGP) is a software tool designed to assist with the analysis of texts following the approach of M.A.K. Halliday. At present, the FGP for English texts runs on any IBM PC AT compatible and is fully integrated with Sprint, a popular word processing package from Borland International. Under a strategic research grant from the City Polytechnic of Hong Kong, work is proceeding toward implementation of a Chinese version of the FGP. In this paper, I will explain the operation of the Functional Grammar Processor as it has been implemented so far for English texts. In addition, I will discuss the proposal for an FGP for Chinese texts.

# The Functional Grammar Processor - a meaning processor

The FGP is a 'meaning processor'; it is not an automatic parser. I would like to draw an analogy between the FGP as a meaning processor and that kind of software we are all so familiar with - a word processor. A word processor assists the writer by making the task of writing easier to accomplish. Of course, the writer must still know what (s)he wants to communicate, the computer does not itself create the document. Similarly, the Functional Grammar Processor assists the user in analyzing text; it also facilitates the subsequent retrieval of information about the text by collecting all the clause analyses into a global database. However, the user must still determine the meaningful elements that make up the text. The computer does not interpret meaning - perhaps that will come later. For now, it only records the interpretation made by the user. The user must know how to analyze the text, the computer does not itself understand the text.

Halliday's approach to analyzing language, called functional grammar, identifies the structures at

clause level that contribute to the meaning of a text. Analysis along these lines shows 'how and why' a text means what it does. Halliday identifies three components of meaning or semantic functions: textual, interpersonal, and ideational. Halliday describes a clause as a complex realization of these three functional-semantic components. Each contributes in its own way to the form of the clause. Corresponding to each component there is some form of structural representation. Textual meaning is represented by theme-rheme and information structures; interpersonal meaning by mood-residue structure; and ideational meaning by transitivity structure.

The FGP includes four separate modules; one for each kind of structural analysis. The modules are independent of one another, each has its own unique terminology and organization. Theme-rheme structure includes elements not found in mood-residue, transitivity or information structures. The same holds true for mood-residue, transitivity and information structures when compared with one another. Each module assembles a partial solution to the larger question of what is the meaning which the text as a whole conveys.

The FGP interfaces between user and text taking one through the process of clause analysis. When the user completes and saves an analysis using the FGP, (s)he is also building a database of entries each of which identifies the elements of theme-rheme, mood-residue, transitivity and information structures that provide the wherewithal to create meaning.

### The look and feel of the FGP

The text to be analyzed is first loaded like any other document into Sprint, the word processor. The example document, TEXT1.SPR, as it appears in Sprint, is shown in Slide 1. The user then highlights the clause to be analyzed using either the keyboard or mouse and chooses a type of analysis from the FGP In Slide 2, the first clause "Stitching pop-up menu. together the ideal computer system for your business can often be a difficult operation" is highlighted. From the Mouse menu, the user selects the Functional Grammar Processor. Up pops the FGP menu; Theme is selected. Next, from the Operation menu, Analyze clause is chosen. With the Functional Grammar Processor activated, the data entry form for Theme-Rheme Structure [Slide 3] now appears on the screen. The highlighted clause is already entered for the user in

the Clause (Complex) field at the top of the form. The topical theme for this clause is "Stitching together the ideal computer system for your business". First we move the cursor to the Ideational field using the tab key, then with every press of the F5 key each successive word is cut from the Clause (Complex) field and pasted to the Ideational field [Slide 4]. Next, the user presses the Return/Enter key and the topical theme from the *Ideational* field is entered in the Theme field [Slide 5]. The contents of the Rheme field are automatically adjusted leaving only the remainder of the clause "can often be a difficult operation." As indicated by the status line at the bottom of the screen, pressing F10 saves the analyzed clause to disk and returns the analyzed clause in a Sprint document. Notice, however, in the lower left hand corner of the screen displayed in Slide 6, that this document is now TEXT1.THM, no longer TEXT1.SPR. Originally there existed only the one Sprint document file, TEXT1.SPR. When the user highlighted the clause to be analyzed and selected Theme-Rheme analysis from the FGP pop-up menu, a new file "TEXT1.THM" was immediately created to receive back the analyzed clause. TEXT1.THM looks exactly like TEXT1.SPR except for that previously highlighted clause which now appears in analyzed form. Once TEXT1.THM has been created, further theme-rheme analysis must be carried out within it and not the original document. In fact, once TEXT1.THM has already been created and you attempt to do theme-rheme analysis on a highlighted clause in the original document, the program will automatically replace the original document, TEXT1.SPR, on the screen with the existing TEXT1.THM file (containing the text view of the theme-rheme structure). Similarly, new text files are created to provide text-views of clause analyses for mood-residue, transitivity and information structures.

### On-line HELP

In order to do transitivity analysis on the same text, the user re-opens the original document, TEXT1.SPR [Slide 7], and this time selects **Process** in the <u>FGP</u> menu. Transitivity analysis involves identifying the process, participant(s) or role(s), and circumstance(s). The possible roles played by participants will depend on the process: actor in a material process, behaver in a behavioural process, carrier in a relational process, etc. The first participant in our highlighted clause is "Stitching together the ideal computer system for your business".

Once the user has entered this participant into the Participant field of the entry form, up pops the Process menu waiting for the user to identify the type of process in which this participant is engaged [Slide 8]. If, however, one is not quite sure what is meant by material process, they can then press F1 and a HELP window appears listing information relevant to the current menu or field. Slide 9 shows the HELP window for Process which lists each process type and its corresponding category meaning, e.g. a material process involves doing, a mental process involves sensing, etc. Press ESCape and the HELP window disappears. The participant is participating in a relational process so we select Relational in the Process menu. Next pops up the Type submenu indicating that there are three types of relational process: intensive, circumstantial and possessive. Again, suppose the user is not sure about the type of relational process to assign to 'can often be' - is it intensive, circumstantial, or possessive? While still in the Type submenu, pressing the HELP key (F1) will bring up the HELP information about the three types of relational process in English [Slide 10]. Because the relationship between the two participants is one of sameness, the user selects Intensive. From the Mode submenu [Slide 11], the user picks Attributive. In attributive mode, the participant can be either the Carrier or an Attribute. Slide 12 shows the user has identified the participant as the carrier. Next the user enters the process into the Process field [Slide Having already identified the first participant as carrier in a relational process, the selection bar in the Process menu is now drawn on Relational. The user presses the Enter key and the Type submenu appears [Slide 14]. The user selects Intensive with the result shown in the Analysis field [Slide 15]. The user next moves the cursor back to the Participant field, presses F6 to clear the field, enters the other nominal group in this clause -- "a difficult operation" [Slide 16], and repeats the process for en-The relational process is Intentering information. sive [Slide 17], Attributive mode [Slide 18], and the element in question is an Attribute [Slide 19]. With the analysis of process and participant(s) now completed [Slide 20] the transitivity structure for this clause is saved to disk and returned to the Sprint document, TEXT1.PPC [Slide 21].

# Viewing an analysis of another structural type

While analyzing a clause for one structure, say mood-

residue, you may wish to see the finished analysis of that same clause for another type of structure, themerheme or transitivity. For example, the user may wish to recall which element (s)he identified as the carrier of an intensive attributive clause. After all, in such clauses, only active voice is possible and the carrier must therefore be the subject of the clause.

As before, the user calls up the FGP from the <u>Mouse</u> menu in Sprint to begin the operation of mood-residue analysis [Slide 22]. In the entry form for Mood-Residue Structure [Slide 23], the user presses the Menu key, F9, and chooses **View Analysis** from the FGP <u>Main</u> menu to view another type of structure for this clause [Slide 24]. The user selects **Process-Participant-Circumstance** from the <u>Analysis</u> submenu [Slide 25] causing the program to search the external database for the requested analysis. Once it is found, the structure is displayed in the *Analysis* field [Slide 26].

Since such information about the transitivity structure is accessible by the user it is also accessible by the program. The program itself without human intervention should be able to recognize the carrier as subject in an intensive attributive clause. This example illustrates the potential then for automating certain steps in the analytical process. Information known about one type of structure already stored in the database should inform the other types.

Continuing with our mood-residue analysis, the user follows the same procedure as before. On-line HELP is available for every field and menu. What is the grammatical subject? [Slide 27] The Subject is the 'resting point' of the argument... What are mood adjuncts? [Slide 28] MOOD ADJUNCTS relate specifically to the meaning of the finite verbal operators... What is the complement? [Slide 29] A COMPLEMENT is an element within the Residue that has the potential of being Subject but is not...

Again, mood-residue analysis once completed [Slide 30] is saved to disk and returned to the Sprint document, TEXT1.MDR [Slide 31].

### Text files as text views

Text-views, those text files to which FGP returns completed analyses, become the focus of interaction between user and text. From them the user can highlight another clause for analysis, or even highlight a previously analyzed clause and modify the previous analysis or delete it.

The text-views also facilitate embedded analyses. For example, the second sentence of TEXT1

But when you have the right connections, everything can be tailor made to suit your needs.

might be analyzed in the following manner at the level of clause complex:

[ref(2), theme([struct(But), clause\_as\_theme(when
you have the right connections)]),
rheme(everything can be tailor-made to suit your
needs.)]

The modifier clause, "when you have the right connections", is in thematic position before the head clause. The analysis does not stop there, however, as we still need to analyze both the modifier and head clauses for theme-rheme structure. In TEXT1.THM, we then highlight just the modifier clause for analysis,

[ref(2), theme([struct(But), clause\_as\_theme(when
you have the right connections)]), rheme
(everything can be tailor-made to suit your needs.
)]

with the result given below:

[ref(2), theme([struct(But),clause\_as\_theme(
[ref(3), theme([struct(when),topical(you)]), rheme
(have the right connections,)])]), rheme
(everything can be tailor-made to suit your needs.
)]

Likewise, the head clause,

[ref(2), theme([struct(But),clause\_as\_theme(
[ref(3), theme([struct(when),topical(you)]), rheme(have the right connections,)])]), rheme(everything can be tailor-made to suit your needs.)]

requires further analysis as shown below:

```
[ref(2), theme([struct(But),clause_as_theme(
[ref(3), theme ([struct(when),topical(you)]),
rheme (have the right connections,)])]), rheme(
[ref(4), theme([topical(everything)]), rheme( can
be tailor-made to suit your needs.)]
)]
```

### Alternate analyses

Where two interpretations of the same clause are possible, one the literal or congruent, the other metaphorical, both analyses can be included. Halliday gives as an example of a grammatical metaphor in the interpersonal component the sentence "I don't believe that pudding ever will be cooked." Here, the opening phrase "I don't believe" functions as an interpersonal (modal) theme:

[ref(3), alternative\_to([1,2]), theme([modal\_adj(I
don't believe), topical(that pudding)]), rheme
(ever will be cooked)]

The list of integer values assigned to "alternative\_to" indicates the clause analysis/analyses where the congruent interpretation of this clause is given:

```
[ref(1), theme([topical(I)]), rheme(don't believe
)]
```

[ref(2), theme([topical(that pudding)]), rheme
(ever will be cooked.)]

# Analyses as terms in a prolog database

Besides the text-views, analyses are also saved as prolog terms in a Prolog external database. External simply means it resides in secondary storage, like on the machine's hard disk. Prolog is what is known as a declarative or database language. For instance, after analyzing the first clause of the document TEXT1 in terms of theme-rheme, mood-residue and transitivity structures, our database would contain the following three terms:

- ~1,[ref(1),theme([e("topical","Stitching together
  the ideal computer system for your business")],
  rheme("can often be a difficult operation.")]
- ~2,[ref(1),mood([m("subject","Stitching together the ideal computer system for your business"), m("finite","can"),m("mood\_adj","often")]), residue ([r("pred","be"), r("complement","a difficult operation.")])]
- 73,[ref(1),participant([e("carrier","Stitching
  together the ideal computer system for your
  business")]), process ([e("intensive","can often
  be")]), participant([e("attrib","a difficult
  operation.")])]

Don't be concerned by the complicated appearance of these terms. The user never really needs to concern him/herself with the internals of the database of terms generated by the FGP. A query program is being developed for use with the FGP to generate reports on the analysis of a particular text. If, on the other hand, you have access to Turbo Prolog, or PDC Prolog as it is now called, and you are interested in designing your own query programs, then a knowledge of the internal structure of the database is essential.

In fact, two external database files are saved for each text, one containing the screen data (what the contents were of each field in the analysis screen when the user saved the analysis and exited to Sprint), the other the analyses themselves. The file containing the screen data is necessary for the functioning of the program - it is not accessible by the user. If, as mentioned before, you highlight a previously analyzed clause, then the analysis screen will appear just as you left it when you saved and This makes it easier for the user exited to Sprint. to modify a previous analysis. The second database file was designed to support further extension of the software in two ways: (a) to enable a user to guery the database for information about consistent patterns of usage - useful if one were looking for instances of foregrounding; and (b) to facilitate future automation of certain steps in the analytical process.

### Plans for future development

Besides the report generator mentioned above, I am presently developing the Functional Grammar Processor to work with its own built in text editor rather than require users to access FGP from Sprint.

### Functional Grammar Processing for Chinese Texts

Implementation of a Chinese version of the FGP will assist in the advancement of research into the similarities and differences between Chinese and English text structures. Such insight should be of interest to scholars in the fields of translation, contrastive linguistics and Chinese linguistics.

Halliday maintains that the ideational, interpersonal, and textual components of meaning are universal to all languages, while their structures - their terminology and organization - are language specific. Thus, while the FGP for texts in Chinese will share a similar conceptual basis with the existing FGP for texts in English, the knowledge modules will require extensive alteration.

The development of the FGP for Chinese texts includes the following steps: (i) identification of field names and structural organization of ideational, textual and interpersonal components of meaning in Chinese written texts; and (ii) implementation of input screens, help windows and menus in Chinese.

The ideational function of language, "expressing the reflective, experiential aspect of meaning" (Halliday, 1985:101) is served by the system of Transitivity which "specifies the different types of process that are recognized in language, and the structures by which they are expressed" (1985:101). A process consists of three components: the process itself, participant(s), and circumstance(s). Halliday suggests that "This tripartite interpretation of processes is what lies behind the grammatical distinction of word classes into verbs, nouns and the rest, a pattern that in some form or other is probably universal among human languages" (1985:102).

## Subject and Topic in Chinese

As regards interpersonal and textual functions, there is obvious contrast in their corresponding structures in English and Chinese. As already mentioned, exten-

sive reworking of the knowledge modules will be required. A basic issue that needs to be dealt with is whether Halliday's approach is indeed suitable for Chinese texts. English grammar is essentially but not exclusively clause-based. As I will point out shortly, Halliday identifies what he calls information units that do not necessarily correspond to clause boundaries. Thus Halliday, while identifying the clause as most basic lexico-grammatical unit, nevertheless recognizes and includes in his functional grammar discourse units other than the clause. This is necessary if comparison and contrast between the two languages is to be made possible within the functional framework of Halliday's approach.

Tsao Feng-Fu in his monograph entitled A Functional Study of Topic in Chinese: The first step towards discourse analysis describes Chinese as a discourse-oriented language in which "discourse elements like topic interact with the syntactic organization of the sentences under their domain" (1979:97). Tsao maintains that both subject and topic can be identified in Chinese, but argues that they be treated as belonging to different levels of grammatical organization. Processes in which subject plays an important role have to do with sentence-internal mechanisms, while processes involving topic extend beyond clause/sentence boundaries.

Tsao identifies the following properties for subject in Chinese (1979:83-84):-

- Subject is always unmarked by preposition.
- In position, subject can be identified as the first animate NP to the left of the verb; otherwise, the NP immediately before the verb.
- Subject always bears some selectional relation to the main verb of a sentence.
- 4. Subject tends to have a specific reference.
- Subject plays an important role in the following co-referential pronominalization or deletion processes: reflexivization, serial verb construction, imperativization, and Equi-NP deletion.

The following properties are given for topic (1979:88):-

- Topic invariably occupies the S-initial posi-1. tion of the first sentence in a topic-chain.
- Topic can optionally be separated from the rest of the sentence in which it overtly oc-2. curs by one of the four pause particles a (ya), ne, me, and ba. Topic is always definite.
- з.
- 4. Topic is a discourse notion; it may, and often does, extend its semantic domain to more than one sentence.
- Topic is in control of the pronominalization 5. or deletion of all the coreferential NP's in a topic chain.
- 6. Topic, except in sentences in which it is also subject, plays no role in such processes as reflexivization, passivization, Equi-NP deletion, verb serialization, and imperativization.

While Tsao's definition of topic identifies it as serving a textual function in Chinese, it clearly cannot be described by reference to thematic structure Halliday distinguishes between Theme-Rheme and Topic-Comment by pointing out that the term topic is often used as a cover term for two functionally distinct concepts, "one being that of Theme and the other being that of Given" (1985:39). An element functioning as Given accompanied by an element functioning as New together constitute what Halliday calls information structure. The distinction between Given and New has to do with whether the information presented by the speaker is recoverable (Given) or not recoverable (New). Whereas thematic structure (Theme-Rheme) is clause-based, information structure (Given-New) may extend beyond a single clause to over one clause and half of the next, or even to two or more clauses. Thus Halliday maintains the existence of information constituents in their own right. units as parison can thus be made between what Tsao calls topic-chains in Chinese and what Halliday identifies as information units in English.

I do not share the reservations of some linguists who regard Halliday's grammar as being inappropriate for analyzing Chinese texts because of its clause-based orientation. According to Tsao there are processes operating at two levels in Chinese, one at the level of clause, the other above the clause. Halliday's approach with its recognition of structures at various levels appears well suited to the task at hand.

### Information structure: analysis beyond the clause

A fourth module has been added to the Functional Grammar Processor for English texts to enable analysis for information structure.

The design of this module as shown in Slide 32 is quite simple. An optional Given element usually precedes the obligatory New. The New information has a phonological realization in that it typically culminates in an element with tonic prominence called the Focus. Owing to this connection between tone group and information unit, analysis for information structure has been largely confined to investigations of spoken discourse, not written texts. However, information structure is also relevant to the analysis written texts. Halliday has this to say about information units in writing (1985:316):

In writing, the principle is that (i) the information unit is a clause, unless some other unit is clearly designated by the punctuation; and (ii) the focus falls at the end of the unit, unless some positive signal to the contrary is given, either by lexical cohesion (no focus on repeated word) or by grammatical structure (predication: it is...that...).

How does one identify the boundary between Given and New in written texts? Much the same way one does in spoken discourse -- the Given refers to something already present in the verbal or non-verbal context.

### Conclusion

The Functional Grammar Processor in its present state compiles a database of clause analyses for texts in English. Add the capability to handle Chinese texts both in terms of clause and discourse structure, begin to automate the analytical process and we have the makings of a system for exploring and later comparing/contrasting the elements, both clausal and discoursal, that together contribute to the textual, ideational and interpersonal meaning of texts in the two languages. Providing greater insight into the structures and their elements in English and Chinese can only serve to advance the prospects for an improved state of the art in machine translation.

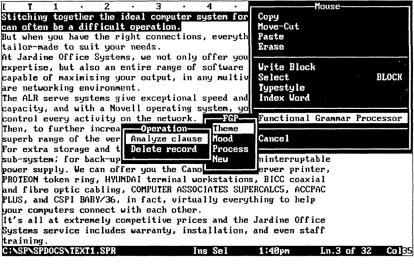
<sup>1</sup> Portions of this paper dealing with the operation of the Functional Grammar Processor were previously presented at the Eighteenth International Systemic Congress, Tokyo, Japan.

### References

- Halliday, M.A.K. 1981. "Text Semantics and Clause Grammar: Some patterns of realization." <u>The</u> <u>seventh LACUS forum</u>, James E. Copeland and P. W. Davis. (Eds.), Columbia SC: Hornbeam Press, pp. 31-60.
- \_\_\_\_\_\_. 1985. <u>Introduction to Functional</u>
  Grammar, London: Edward Arnold.
- Tsao, Feng-Fu. 1979. <u>A Functional Study of Topic in Chinese: The First Step Towards Discourse Analysis</u>. Taipei, Taiwan: Student Book Company, Ltd.
- Webster, Jonathan J. 1990. Presented the Featured Lecture at the 17th LACUS Forum held August 1990 at the State University of California, Fullerton. Title of lecture: "Linguistic Information Processing using the Functional Grammar Processor". Publication forthcoming.
- \_\_\_\_\_. 1991. "Text Analysis using the Functional Grammar Processor" paper presented at the CPHK & HKUST Seminar on Text Analysis in CALL. Publication forthcoming.

```
3
                                                                  11.
Stitching together the ideal computer system for your business
can often be a difficult operation.
But when you have the right connections, everything can be
tailor-made to suit your needs.
At Jardine Office Systems, we not only offer you unrivalled
expertise, but also an entire range of software and hardware
capable of maximising your output, in any multivendor or local
are networking environment.
The ALR serve systems give exceptional speed and performance
capacity, and with a Novell operating system, you can monitor and
control every activity on the network.
Then, to further increase your capabilities, we can offer you a
superb range of the very latest equipment.
For extra storage and tape backup capability, there's the CORE
sub-system; for back-up power, we have the ELGAR uninterruptable
power supply. We can offer you the Canon LBP-811 server printer,
PROTEON token ring, HYUNDAI terminal workstations, BICC coaxial
and fibre optic cabling, COMPUTER ASSOCIATES SUPERCALCS, ACCPAC
PLUS, and CSPI BABY/36, in fact, virtually everything to help
your computers connect with each other.
It's all at extremely competitive prices and the Jardine Office
Systems service includes warranty, installation, and even staff
training.
C:\SP\SPDOCS\TEXT1.SPR
                                                   1:39pm
                                                               Ln.2 of 32
                                                                             Col3
                                      Ins
```

[ Slide 1 : Original text document in Sprint ]



[ Slide 2 : Calling FGP from Mouse menu in Sprint ]

Clause as theme					
Textual————————————————————————————————————		structural		conjunct	ive adjunct
-Interpersonal mode		al(Adjunct) finite		erb	WH-(?)
heme:					
	-	w Alm Hamilton		on Con un	ur business can

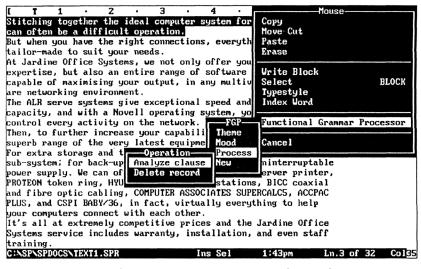
[ Slide 3 : Theme-Rheme Structure Entry Form ]

	[ Theme-Rheme	Structure 1-				
Clause (Complex): can often be a difficult operation.						
Clause as theme						
Textual continuative	structural	conjur	nctive adjunct			
-Interpersonal vocative	modal(Adjunct) finite verb		WH-(?)			
Ideational Stitching together	the ideal computer	system for your bu	siness			
Theme:						
Rheme: Stitching to	ogether the ideal co	mputer system for	your business can o			
Reference No.	1 Alt	Document C:\SP\SPD	OCSNTEXT1			
scl-Exit [F1]-Help	[F5]-Write [F6]-Eras	e [F7/F8]-Shift [F	9]-Menu [F10]-Save			

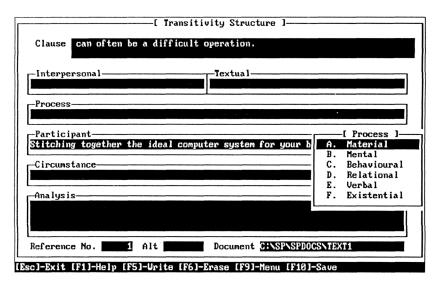
[ Slide 4 : Topical Theme identified ]

\_\_\_\_\_I Theme-Rheme Structure ]\_\_\_\_

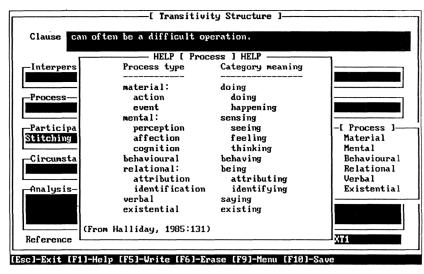
Clause (Complex): can often be a difficult operation.							
Clause as theme							
Textual continuative structural conjunctive adjunct							
Interpersonal vocative modal(Adjunct) finite verb WH-(?)							
Ideational Stitching together the ideal computer system for your business							
Theme: topical(Stitching together the ideal computer system for your busines							
Rheme: can often be a difficult operation.							
Reference No. 1 Alt Document C:\SP\SPDOCS\TEXT1  [Esc]-Exit [F1]-Help [F5]-Write [F6]-Erase [F7/F8]-Shift [F9]-Menu [F10]-Save							
[ Slide 5 : Theme-Rheme analysis completed ]							
[ I							



[ Slide 7 : Calling FGP from Mouse menu in Sprint ]



[ Slide 8 : Participant in what type of process? ]



[ Slide 9 : On-line HELP -- Process ]

Clause can often be a difficult operation.					
-Interpersonal	-Textual				
There are three types of relational process in English:  (i) intensive - the relationship i one of sameness; the one 'is' the oth (ii) circumstantial - the relations between the two terms is one of time, place, manner, cause, accompaniment, matter or role.  (iii) possessive - the relationship between the two terms is one of owner ship; one entity possesses the other.  (From Halliday, 1985:114,119,121)	Intensive Circumstantial Possessive Behavioural Relational Verbal Existential  nt C:\SP\SPDOGS\TEXT1				

[ Slide 10 : On-line HELP -- Types of Relational Process ]