

Startup and Basic Reference Guide to the Multilinear Discourse Analysis Software

Phil Quick

MDA 1.0 (Beta 4.1 -- March 2006)

1. INTRODUCTION*

This short manual will serve as a startup and basic reference guide until the more comprehensive MDA guide is completed. This manual helps you get started with your own language data and with the sample language data provided. For some expanded information, see the Beta User's Reference Guide document and/or review the powerpoint introduction to MDA.

2. A QUICK OVERVIEW OF MDA

2.1 Installation

The MDA program is available on a CD (or downloaded from a web site to be announced). MDA will run on Windows 3.x, Windows 9x, and XP. Find the setup.exe file using file manager. Then double click on this file and the installation process installs the software on your computer (see the readme.1st file for additional information and information on a few bugs). The installation process will provide you with an opportunity to designate the drive and directory if you want a different location than the default.

After you have the MDA icon available, then you can run the MDA program (you can add it to your desktop as a shortcut icon). At the top of the MDA program you will see the menu bar with different menus. These menus will be summarized in the sections below.

See the README.1st file for any last minute changes. The CD may also have additional default file settings and documentation that can be copied to your computer without reinstalling the whole MDA package if you need the original file settings. There will also be zipped or Word 6 Doc files that will have a copy of this reference guide and the MDA Demo B+ paper presented at the 1996 CARLA conference (referred to several times in this paper).

2.2 Features and Development Background

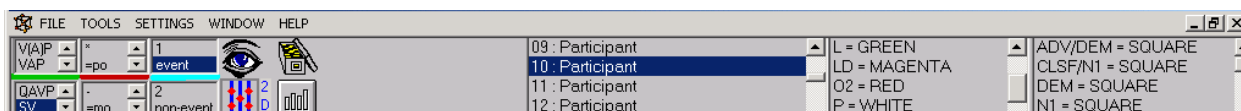
MDA was first developed in Visual Basic 3.0 as a very minimal program called Demo A. The first significant demo version was developed under Visual Basic 4.0 (MDA is currently only in 16 Bit) and was called Demo B+. Kevin Madden joined me as a partner by contributing the original online helps (we also anticipate his increasing participation in MDA version 2 when he will begin helping to develop a three-dimensional Map display). This was the version that was demonstrated at the SIL 1996 CARLA conference in Waxhaw, North Carolina, USA. The paper describing the demo B+ version was printed in the conference proceedings with the title "The Multilinear Discourse Analysis Software Demonstration". Some parts of this paper are already obsolete, however Beta users should look at this paper as a resource whenever this guide fails to provide the information you are looking for.

MDA is primarily designed for advanced linguistic analyses of languages. In order to use MDA, the analyst should have a good grasp of the language he/she is analyzing. Beginners can use this program to learn some linguistic concepts by using the sample data provided, which has the settings already provided.

* I want to personally thank participants at the 1996 CARLA conference for their comments on the Demo B+ version. Also I want to thank Russ Cooper, Wayan Pastika, and David Andersen. The latter two have helped to contribute data and analyze Balinese and Hebrew following the MDA coding parameters, their interaction has been of great value to the development of the MDA program. I also need to mention some help in programming expertise on a couple of programming problems I have had. Alistair George and Marten Gans have contributed ideas and coding methods that resolved some problems I was having.

2.3 The Menu Bar

Across the top of MDA the Menu Bar has menu options categorized under these titles: **File, Tools, Settings, Window, Help**. You can click on any of these and a menu will drop down giving you choices to select from. The **File** menus contains options pertinent to printing, opening a 2D Map display, exiting the program, and viewing contents of other files. **Tools** contains the primary functional discourse tools that the linguist will utilize. The **Tools** include making the MDA database from a shoebox database, as well as the quantitative analytical tools based on the MDA databases. **Settings** are generally the settings that the linguist will need to setup for a particular language that he is analyzing. Many of these settings are critical to properly analyze your language. This will be the most difficult part of the MDA program to deal with.



The **Windows** menu option provides Windows users with the typical tiling capabilities, and allows simultaneous tools open at the same time. The **Help** menu option lists various helps, some of these are typical dialogue boxes on the basic background of the program. One option provides access to online help. The new online help feature uses your Web Browser to view this documentation. Older documentation for the Demo B+ version is accessible with the F1 key.

2.4 Steps to Analysis with MDA

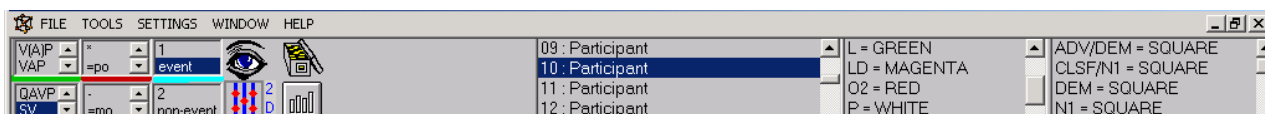
The first step is to annotate a text that you want to analyze. This is the most critical step. The linguist has to carefully study the tagging system that MDA utilizes. This tagging system then must be applied manually by the linguist to each clause of the text. MDA has been designed so that users who are familiar with using Shoebox can just add a few extra annotated fields to the interlinearized text. They can either do this when they are interlinearizing or return to the text after it has been interlinearized.

The second step is to run MDA and set the settings. Then you run the Shoebox tool to make the MDA database (which uses the extension *.MDA, or something like *.MLT is better as *.MDA is reserved by the Windows operating system for other functions, although this will work as *.MDA).

The third step then is to make use of the MDA database either in the graphic 2D Map view or to quantify the text you are analyzing.

2.5 The Toolbar

The bar below the menu bar is the toolbar. The toolbar has 4 icons and 9 list boxes. All of the icons are shortcuts to the main tools the MDA user will use. The 'eye' icon opens a window that allows the user to view files produced by MDA or other files (but it is limited to 32 K). The Shoebox icon is a shortcut to the option on the tool menu called 'Shoebox'. This is the command for making an MDA database from a Shoebox database (or other ASCII file using the MDA tagging system) that has the MDA tagging conventions. The icon which has blue lines and five red dots with '2D' placed vertically beside the icon accesses the 2D Map graphic viewing of the MDA database file. This is for participant tracking and span analysis work. The fourth icon which has a bar graph accesses the quantification part of MDA. This is the part of the program which allows the user to get topic continuity statistics and other statistics based on the MDA database file.



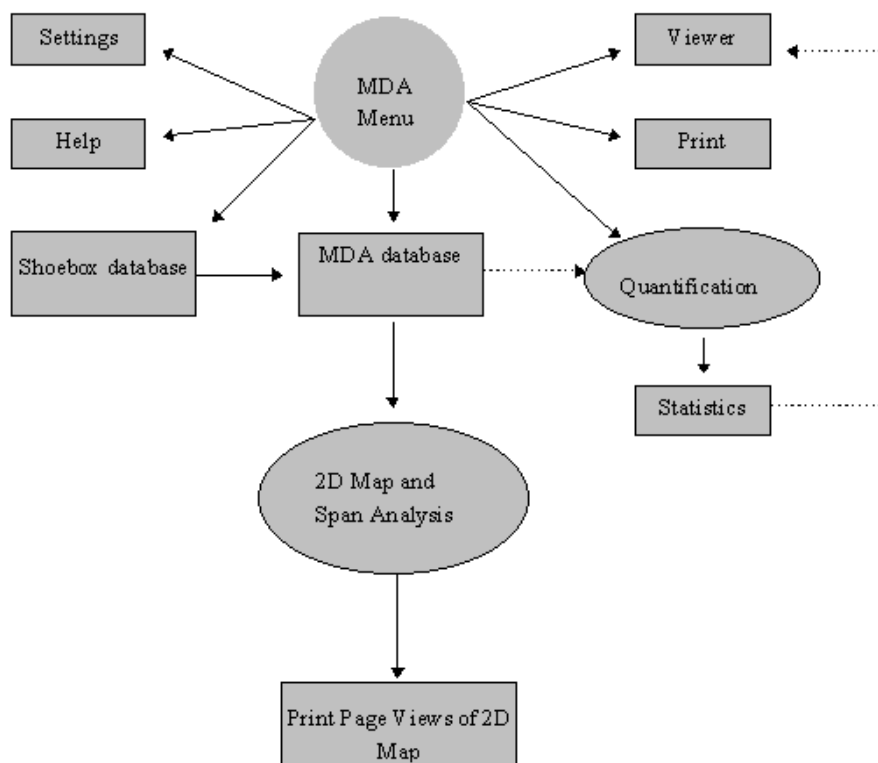
The big listbox in the middle of the toolbar is for helping the user keep track of the participants that are used in the text currently being viewed in the 2D Map view. The participants are entered in one of the setting options -- under tracking options (the middle tab called participant names).

The two smaller list boxes to the right of the participant listbox each keep a current list of the user's codes. These are another reference for the user to quickly check which settings he or she has set under the tracking options. One list box shows which colors the user has selected for grammatical relations or grammatical roles. The other list box shows the shapes associated with different NP types. The six list boxes on the left are in three pairs and correlate to the three pairs of span analyses (the colors between the list boxes indicate which span pairs are being traced).

The icons and list box helps are available throughout the program.

2.6 MDA Flow Diagram

This shows a flow diagram of the main features and components of the MDA program.



2.7 Changes from the Demo B+ Version

The biggest change in the Beta version is greater flexibility for the user. There are now only three full screen windows: The file previewer ('eye' icon), the 2D Map viewing, and the basic window view. The Shoebox database tool options for making the MDA database is now relegated to a dialogue box. The quantification tools are now options available as choices in a dialogue box. The icon of a man and woman and its menu choice for outputting a primitive ASCII version of the span analysis has been removed from the Beta version. This has been superseded by printing options that are almost WYSIWIG with the 2D Map view.

2.8 Bugs and Problems

Printing a hardcopy still has some problems with correct graphical representation. These are primarily color and color fill in problems. It will still be 'good enough' for analysis though.

Some dialogue boxes may only be useful with the mouse (i.e. the enter key, escape key, and double clicking may not have been implemented in all of these places yet).

There are occasional times when an error message may come up--if so the user should click OK and still be able to continue your work. Please report any problems or bugs to: Phil.Quick@sil.org

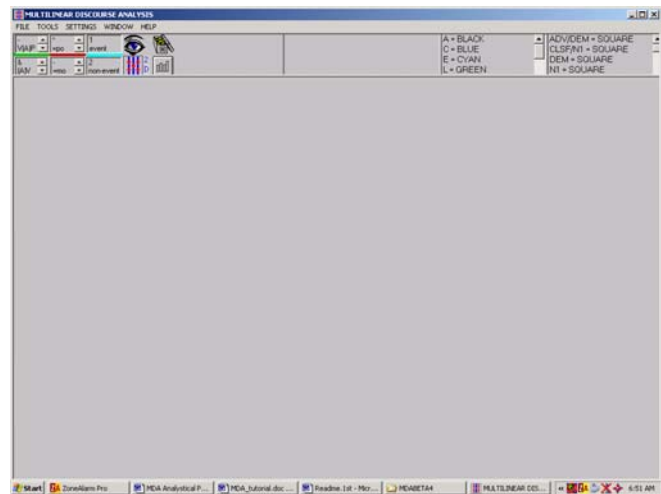
2.9 Database Tagging Construction Rationale

The tagging convention used in annotating the Shoebox database is intentionally split into two types. The first type include the typical fields which contain a range of information for one type. These include fields such as word order type and other discourse information type that the use wants to trace simultaneously with the participant tracking. This information is strictly used and represented in the span analysis (the area of the 2D map referred to as the *slate* is on the left hand side of the screen). The second type of information is more complex and contains up to five pieces of information about an individual participant (or noun phrase) of a particular clause. Technically this information could be spread out into five separate fields of their own. However there are several reasons that these five domains are coded into one field. First it is easier for quantification purposes, secondly in Shoebox it is easier to select from a list one time rather than five separate occasions.

Thirdly, and this is the main reason, there are theoretical aims behind the development of this program that suggest that for development of artificial intelligence this may be the best solution.

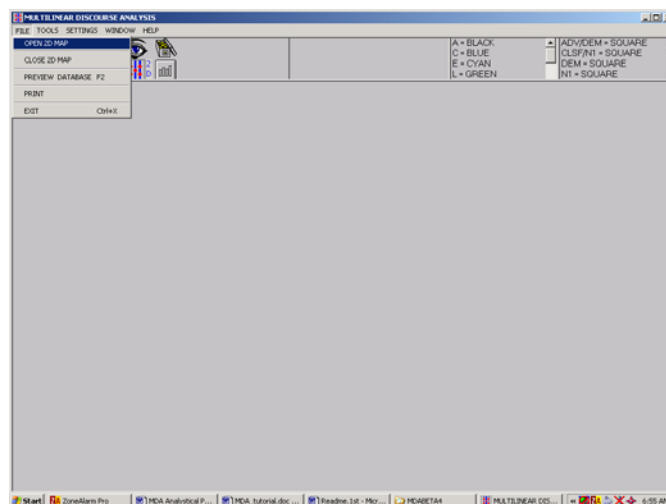
2.10 A Tutorial Lesson

These lessons are aimed at giving a short series of easy instructions on learning to use MDA from the beginning, step by step. For an overview of the MDA program, its features and capabilities, I recommend the user review the powerpoint program first.



2.10.1 Lesson one—Becoming familiar with the 2D map¹

In this lesson the user will become familiar with the main visual features of the 2D Map. To do this you will use one of the sample MDA database files. After installing the program, locate the MDA application file name **MDA_B4_A.exe** if its not already on your desktop (if not yet on your desktop you can create a shortcut icon there). The default folder name is **MDABETA4**. Double click to run the program. You should now see something similar to the screen capture (to the right). Most of the program below the tool bar will appear blank when you first open the program. The next step is to open the menu **FILE** and choose option **OPEN 2DMAP**. This is shown in the next screen capture (on your right here).



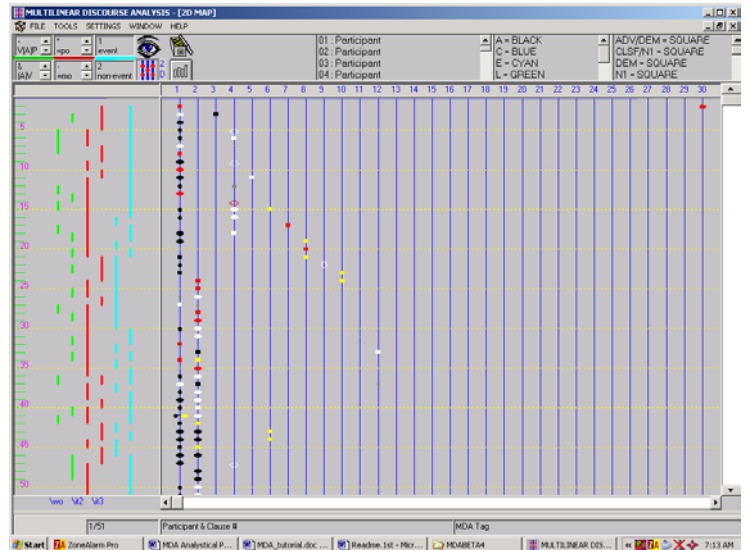
After clicking on **OPEN 2D Map** option you will be asked to select a file. This dialogue box will show a view of files and folders (see screen captures below). You will also notice that the background has loaded an empty display of the 2D Map, and in a few moments we'll discuss

¹ Note that this lesson assumes that none of the files have been modified. If the user has already explored some of the sample files, some of the setting files may have changed. In order to restore the settings to their initial default settings, copy the Pendau.ini file from the **INIBACKUP** folder to the **MDABETA4** folder.

Startup and Basic Reference Guide to the Multilinear Discourse Analysis Software 1.0 (Beta 4.1)

First move click and hold the left mouse button down and move the mouse from left to right while you watch the list box with the list of participant names.² You will see that as the cross-hairs of the mouse change from one participant to another that the listbox is continually updated to match the participant.

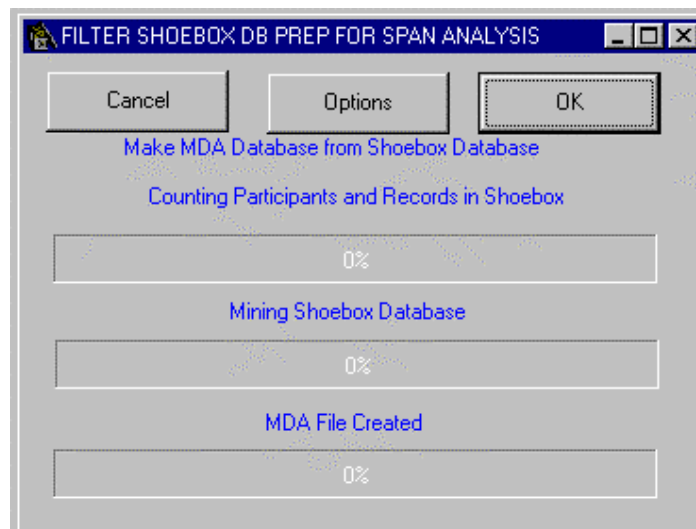
Finally, the last step in this lesson is to close the 2D Map. To do this choose the **FILE** option and select the menu option **CLOSE 2D MAP**. The visual interactive area will close, but the program will remain active to return to the 2D Map or to use other features. To exit the MDA program, choose **EXIT** from the **FILE** menu options.



3. TOOLS

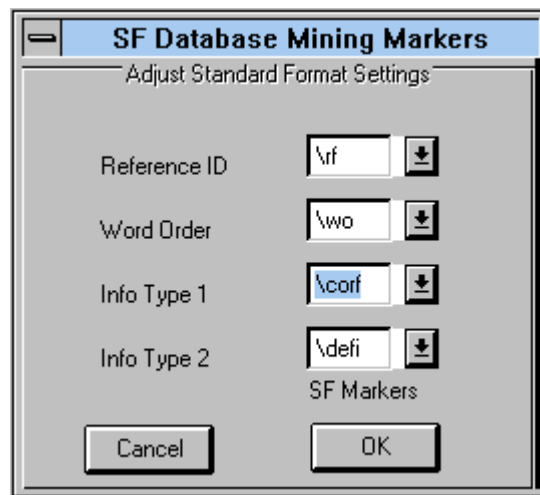
3.1 Shoebox Database Conversion to MDA Database Tool

This is the first option in the **TOOLS**, labeled *Shoebox* and is also accessible with the shoebox icon. This is usually the first step in processing a Shoebox database. The user will need to adjust the settings that are used in the conversion process. This can be accessed from either the **SETTINGS** option called *SF Database Mining Markers* (where SF is the abbreviation for SIL's standard format marking system), or accessed directly from the the *Shoebox* dialogue button labeled *Options*. Status bars alert the user to progress of the Shoebox database conversion to the MDA database.



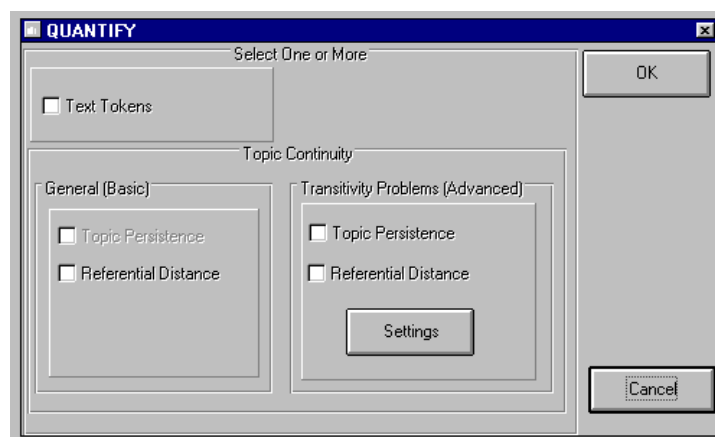
The *SF Database Mining Markers* dialogue box has four combo boxes for the user to set. These settings must be set to match the record and field markers used in the Shoebox database (or non-shoebox database). These settings and this dialogue box is explained in more detail under the section on **SETTINGS**.

² These are listed generically as Participant 1, Participant 2 (and so on) in the initial loading of a text. Note however that the user has the option to specify the names of participants. This will be discussed in a later lesson.

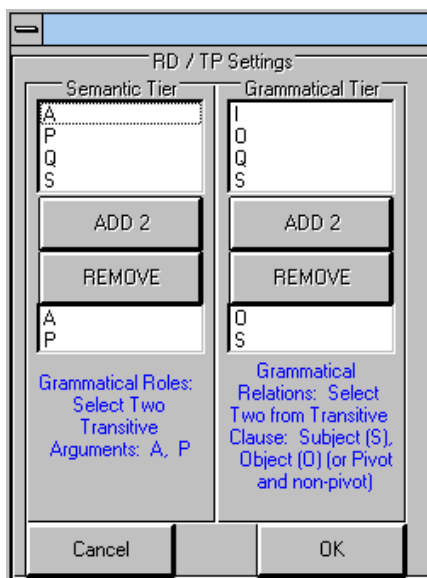


3.2 Quantify Dialogue Box

The *Quantify* dialogue box is accessed under the **TOOLS** menu option or with the graph icon on the tool bar. The *Quantify* dialogue currently allows the user to choose from one to four statistical outputs (either separately or joined together in any combination). There are currently two types of Topic Continuity statistics available: 1) general or basic topic continuity, and 2) advanced topic continuity (transitivity problems). Basic topic continuity analysis examines the referential distance for any NP type and includes analysis for both grammatical relations and semantic roles. For advanced transitivity problems, the two options are topic continuity statistics for *Topic Persistence* and for *Referential Distance*. These include Givón's basic methodology and Dryer's supplementary method. Using the Givón approach one table of data is created, and using the Dryer approach there are two tables of data created. Altogether, the topic continuity options yields seven different analyses. The third quantification option produces statistics on all of the noun phrases, and is labeled as *Text Tokens*.



The *Quantify* dialogue box also has a button labeled as *Settings*. This accesses the *RD/TP Settings* dialogue box (this can also be accessed under the menu options labeled as **SETTINGS**). The user must enter the proper settings in this dialogue box in order to get the proper Topic Continuity statistics. This copies two of the tier settings that the user sets in the **SETTINGS** option labeled *MDA Tag Positions and Codes* (see below for more details). The user will need to select only two correct abbreviations from each tier which correspond to the two primary grammatical roles (from the *Semantic Tier*), and the two primary grammatical relations (from the *Grammatical Tier*).



The user must consult the *required readings* I've listed at the end of this paper in order to properly understand the topic continuity statistics. Especially note that the analyses called "vertical" and "horizontal" analyses is only found in Dryer's article (1994; I do not explain these statistics at all here).

3.2.1 Topic Persistence

The algorithms used in the topic persistence follow the revised procedures in Givón's 1994 book *Voice and Inversion*. The user should however be aware that the computer stops counting 10 clauses from the end of the text (this is partly an arbitrary decision and partly a necessity to implement the counting methodology). [Note: The zero-th clause has not yet been implemented in the statistics--but I plan to put this in.]

3.2.2 Referential Distance

The algorithms used in the referential distance also follow the revised procedures in Givón's 1994 book on *Voice and Inversion*. The user should be aware that the program starts counting at the fifth clause from the beginning (this is required by the computational methodology I have employed).

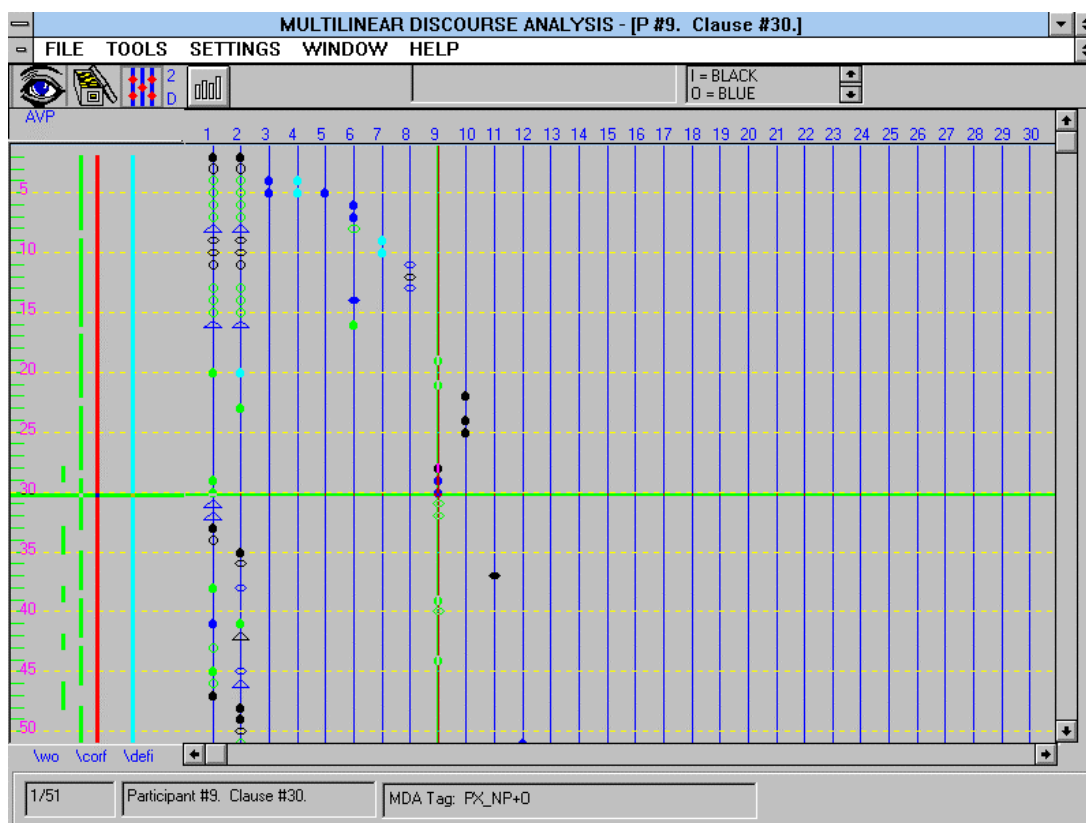
3.2.3 Text Tokens

The text tokens option summarizes the total of all noun phrase types that the user has identified in the Noun ID tier (this gives a profile for an individual text). The summary shows the statistics of every noun phrase type and its totals for every participant. The user should be aware that there may be some imprecision here depending on how the analyst has tagged the database. If the user identifies plural entities coded as a single pronoun or collective noun as separate instances, then I believe that each of these instances will be counted, and will therefore add more to the totals than there really is (I will have to reflect on whether the program will be able to fix this problem).

3.3 Span Analysis and NP Tracking: 2D Map

This can be accessed with the option labeled as *Span Analysis and NP Tracking: 2D Map*. This is arguably the most interesting part of the MDA program. This option allows the user to visually represent his or her text's database and has the most features built into any one aspect of the total MDA program. The text is represented in a grid of 30 participants by 50 clauses in any one screen view. Scroll bars allow the user to scroll past these boundaries up to the limits of the database (the Beta version is not limited to 30 participants and 360 clauses as the Demo B+ version was).³

³ The beta 4.1 version limits participants to 95 participants.



In order to help the user refer to different sections of the 2D Map these are differentiated with two terms. The main part of the map is the participant tracking section which can be referred to as the *Map*. The map is the part of the screen view that shows participant tracking. Blue vertical lines numbered 1 to 30 (in the default screen view) appear with horizontal dotted yellow lines spaced at five clause intervals. At the intersection of the clause and participant a graphic symbol locates the participant on the grid or map. This symbol can be set using colors and shapes to visually allow the user to trace the noun phrase type and either grammatical roles or grammatical relations within the entire context of the text. The area to the left of the map has the clauses numbered every five clauses where the yellow dated line intersects. Each clause position is identified by a short green tick or line. This area between the clause numbers and the participant map can be referred to as the *slate*.

The slate is the area of the total 2D Map where the span analyses are traced. The span analyses contrasts pairs of items listed in the **SETTINGS** option labeled as **2D Map: Slate Span Group Settings**. Typically one of the spans that is traced will be contrasting to different word orders. Each Span Group (word order would be one group) uses the same color in a binary contrast. Currently one pair is green, the middle pair is red, and the third pairing is cyan.

At the top of the 2D Map are three list boxes that are used as keys to remind the user what settings he/she has used and which number matches which participant. The participant settings are set in the **SETTINGS** option labeled **Tracking Options**. The **Tracking Options** have three tabs to choose from, and the middle tab is labeled as **Participant Names**.

Clicking on the left mouse button and holding it down shows cross-hairs that allow the user to identify the MDA code for a particular participant and allows the user to visually see which clause and participant the mouse is pointed at. At the bottom of the 2D map is a status bar which identifies the current or last position clicked on with the mouse, where the position is the clause number and the participant number. When the participant names have been set, the participant clicked on is focused on automatically in the list box on the tool bar. The small text box on the far left on the status bar at the bottom of the screen identifies the first and last clause that is in the current screen view. The first and last clause are divided with a division line (e.g. 1/50 means clauses 1 to 50 are currently being viewed). The third text box used on the bottom of the status

bar provides the entire MDA code of one or two participants for the clause and participant that has been clicked on with the mouse.

Other useful information is shown above and below the slate area of the 2D map. At the bottom of the slate each of the slate groups is identified with the field markers used in the database. Whenever the user clicks on a clause the MDA information for that specific clause will appear at the top of the slate until the user clicks on a different clause. This way the user can see whether the clause is SVO or VOS, etc, and helps to know which spans are being traced in a particular clause or series of clauses.

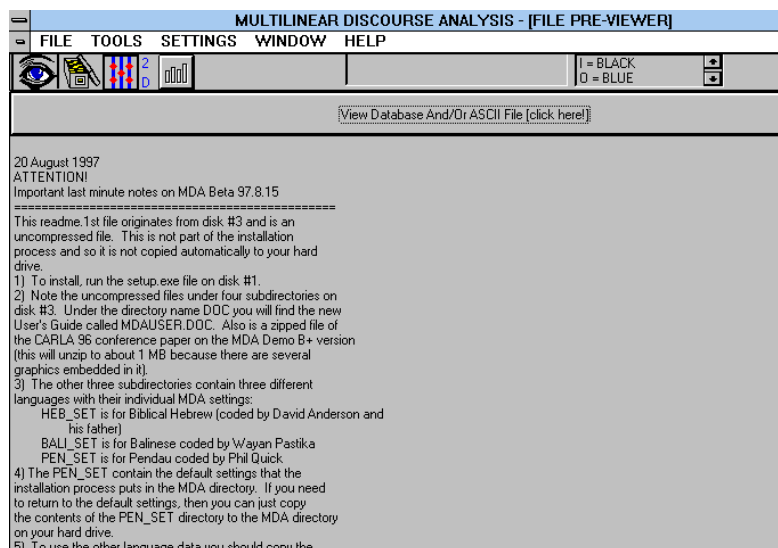
When the user clicks on the vertical scroll bar the clauses are incremented by 20 clauses in either direction, depending if it is above the scroll tab or below the scroll tab. The horizontal scroll bar increments the participants by five participants at a time. The scroll tab for either the horizontal scroll bar or the vertical scroll bar can be grabbed by the mouse by clicking on the mouse and holding it down and moving it to the approximate location you want the map to move to. The arrows on the scroll bars increment participants by one and the clauses by one.

Multiple instances of the same participant can be tagged in the database by repeating the same field marker and by filling in the differences accordingly. Two instances of the same participant are shown graphically with a short horizontal blue line with the shape and color at the end of both sides of the blue line. The user may notice that an instance of one of the two participants is a dark gray color, this means that the program has identified the multiple instances, but that one of the instances has not been abbreviated properly (note that the dark gray color is not one of the color choices available to the user).

4. FILE

4.1 Preview Database

Opening up the *Preview Database* option shows a full screen size text box. This can also be accessed by clicking on the 'eye' icon. Between the text box and the tool bar is a screen-wide button labeled *View Database And/Or ASCII File [click here!]*. When the user clicks this button a dialogue box appears for the user to select the file he or she wants to view in the text box. This text box is restricted to about 32 K (and sometimes up to about 64K) because of the programming language constraints. The main purpose of this textbox, however, is to show the results of the Quantification output. When the user selects any one or more of the Quantification choices the output is automatically loaded into the Preview textbox. The user is also asked to name a file for the output to also be output to a file. This gives the user the choice of opening up the Quantification output in a word processor, or to select the information in the textbox and cut and paste it directly into your word processor file.



When you already have information in the text box and you click the View Database button, the text box is cleared to allow the user to open up a new file to be viewed. This textbox does not provide a means for editing, it is simply a viewer.

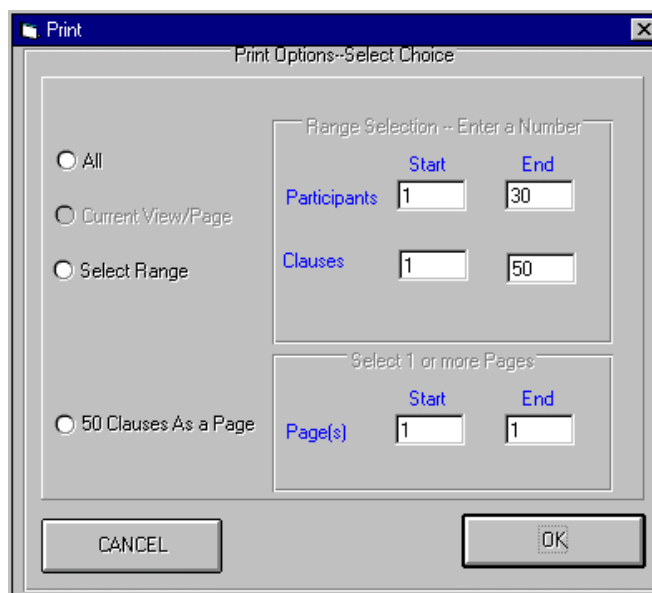
To close the text viewer the user needs to use the Windows procedure to close the window inside of the program window (for Windows 3.1 this is the small hyphen in the upper left hand corner, clicking on this opens a menu--choose close and then the text viewer will be closed; Windows 95/98 uses a similar procedure but has a miniature icon in the upper left hand corner which you click on).

4.2 Open or Close 2D Map

Currently only one graphic view of the 2D Map can be run at one time. In order to open a new 2D Map view click on **Open 2D Map**, or to close the current 2D Map display click on **Close 2D Map** under the **File** menu option. This does not exit the MDA program, it only closes the view of the 2D Map. Once it is closed a new file can be selected to view as the 2D Map by clicking on the 2D Map icon or selecting the 2D Map choice under **Tools**.

4.3 Print

The **Print** dialogue box is currently only accessible from the **FILE** menu option. This dialogue box allows the user to print a replication of the 2D Map graphic view of participant tracking and span analysis of a text. The **Print** dialogue box gives several options [*the Beta 4.1 version disables other options currently—only the **all** option is possible currently*]. Selecting the **All** option prints out the entire 2D Map (but in the Beta version may only include the first 30 participants). The second option **Select Range** allows the user to select a range of clauses and participants (in the Beta version it will print out a full page on the last 'page'). The **Range Selection** is only available when the **Select Range** is activated or checked. The third option dialoug box allows the user to print out units of 50 clauses as if they were a 'page' (i.e. equivalent to a screen view).



Clicking the **OK** button takes the user to the dialogue box to select the MDA file the user wants to print from. The 2D Map view does not have to be opened in order to print a selection. Currently the printing is not linked to the file currently being viewed, these work independently of each other in the Beta version.

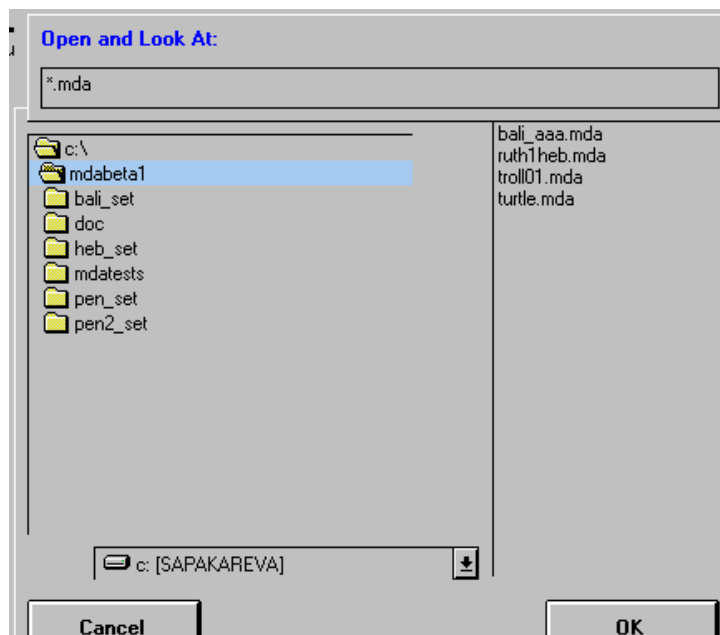
Clicking the **Cancel** button exits the **Print** options dialogue box.

4.4 Exit

Click on *Exit* to exit the MDA program. Any settings that have been modified have already been saved if the user has clicked OK at the settings' dialogue boxes, otherwise nothing new will be saved, i.e. there isn't any dialogue box asking if you're sure you want to do this.

5. THE FILE SELECTION DIALOGUE BOX

This dialogue box is not directly accessed. The *File Selection Dialogue Box* (this dialogue has no title), allows the user to locate the file he or she wants to open or process, depending on which option is being used. This allows the user to change drives, change directories, use wild cards to view specific file types just as you can in other windows programs. A short message in blue appears above the text box in which the file you select will be placed. Clicking on the *OK* button confirms that is the file the user wants, and proceeds to process or open the file selected. This action can also proceed by double clicking on the file name in the list box where it appears with the other file names.



Clicking on the *Cancel* button exits the dialogue box.

6. HELP

6.1 Online Help

The *Online Help* option opens up your HTML browser and loads a simple HTML version of similar to this manual (but more like the MDA User's Guide).

Online help of the previous demo versions (including Beta 1 released in 1997) can be accessed by pressing the *F1* function key. The first helps section is called *MDA Software Helps*. The second helps section is a preliminary introduction to the Multilinear Discourse Theory, with the helps jump labeled as *Multilinear Discourse Theory Helps*. The Multilinear Discourse Theory has not been fully developed, but it is the backbone of the MDA software and is the basis for the tagging system that the user needs to use, however the MDA software is flexible enough to allow other theoretical views to be implemented within the tagging system [there are some theoretical issues and problems that have not yet been resolved, but these will not be discussed in this reference guide].

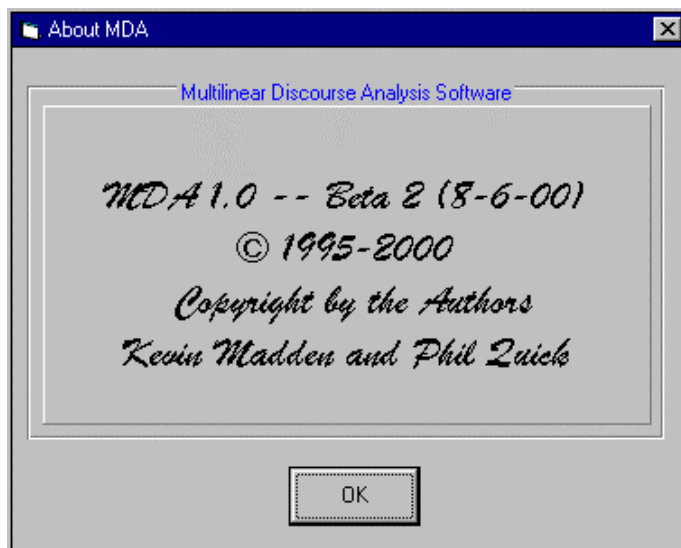
[NOTE: Release 1.0 will use HTML, and this document will be accessible from within MDA as an online help, but it will be called via your current HTML browser automatically; for now this document is available as a non-hierarchical HTML document

6.2 What is MDA?

This is a simple information dialogue box that has been with the MDA program since the first Demo A version. This will probably be removed when MDA 1.0 is released.

6.3 About MDA!

The *About MDA* dialogue box provides information on the version number of the MDA program, the name of the software and the authors of the program. Click *OK* to exit the dialogue box.



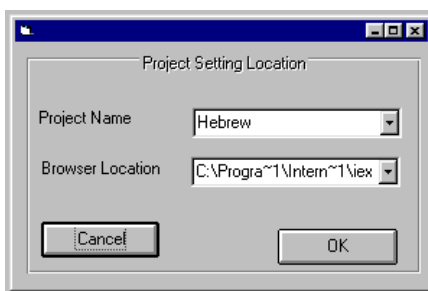
7. SETTINGS

MDA settings are essential in providing flexibility for linguists in describing a wide variety of languages and a wide variety of abbreviations according to the analyst's taste. The new user should work through all of the settings when preparing a new language to analyze. I suggest that all Beta users should look at the sample databases provided and try the various procedures out on one or more of the sample databases, before attempting the major task of developing codes for the language you will analyze. This would be the best training you can get to prepare you for coding the language you analyze. I also suggest starting with only 20 clauses (and it has to be at least 20 clauses, not less in order to run some of the statistics procedures in the Quantification).

All settings are stored by the program so that the user does not need to enter these over and over again. This also allows the user an easy way to modify any abbreviations he or she may change in the source Shoebox database (or non-Shoebox database).

7.1 Project Settings' Location

The *Project Setting Location* dialogue box allows the user to designate the current project name. This will create a main *ini* file which will reference a project specific *ini* file which will store all of the users choices and settings. Different languages should have different project names. Significantly different ways to analyze one language can be given different project names as well. Selecting the *OK* button stores the user's choice for the project's name.

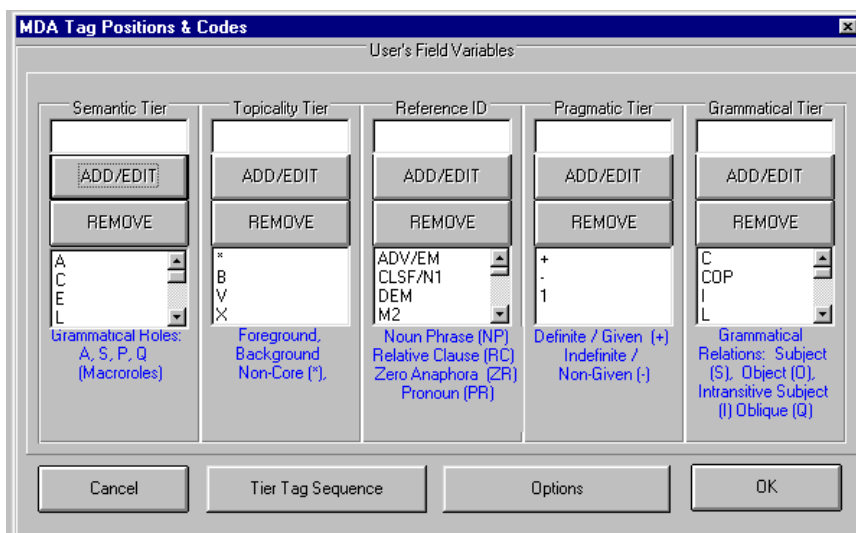


The browser location allows the user to select their web browser of choice to use as the online helps [new update: note that this is no longer necessary as of Beta 4.1, the browser will be launched automatically from the helps menu]. The MDA installation program creates a subdirectory called *mda_html* which has all of the html documentation and graphics to use with the online help. This directory should not be moved, otherwise MDA will not be able to locate these files (although the help can still be accessed manually by launching your web browser independent from the MDA program and finding the appropriate html files manually--just like browsing offline)..

Selecting the *Cancel* button leaves the project and browser setting as they were.

7.2 MDA Tag Positions and Codes

The MDA Codes and Tag Positions is the most important part of the MDA program to be set properly. The annotations used in the Shoebox database are separated into *tiers*. Each tier has its own set of abbreviations. For every participant in a clause, each of these tiers has to be symbolized by an abbreviation in order for the MDA program to recognize the participant as a valid item (otherwise it will ignore that participant). If an abbreviation is left out, the program will ignore all participants with that abbreviation.



The user may want to follow the dialogue box in the order of which the abbreviations are tagged in the database, however this is not a requirement. The only requirement is that the user must always use the same sequence, and must indicate which sequence he or she is using (this is in the button on the left next to the *Cancel* button marked *Tier Tag Sequence*, which is explained below).

The kinds of information that fill in each *tier* have been explained in section 4.2 of the 1996 paper (Quick 1996), and the Beta user should have that on hand. Below each of the *tiers* there is a brief note suggesting abbreviations and a concise statement of the kind of information that belongs to each *tier*.

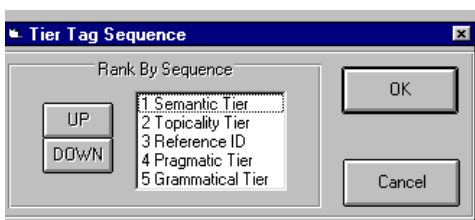
Each *tier* has an upper box for entering new abbreviations, and a lower box that maintains a list of the abbreviations the user stores in the settings file for that *tier*. There are two buttons for for entering and removing abbreviations for each *tier*. To enter a new abbreviation type in the upper box and then click on the *ADD/EDIT* button. To remove an abbreviation from the list box, click on the abbreviation that you want

to remove. This should now be highlighted. Next click the *REMOVE* button and the abbreviation is removed from the list.

Click on the *OK* button to save and store the new settings. Click on the *Cancel* button to close the dialogue box and ignore any changes you may have made.

7.2.1 Tier Tag Sequence

The *Tier Tag Sequence* option button accesses the dialogue box labeled **Tier Tag Sequence**. This dialogue box shows a list box of the sequence of the tiers. Since different users may have different preferences on sequence the abbreviations, this allows the user to set their own sequence within the parameters of the MDA program. The user should only be interested in adjusting the *tiers*' sequencing to match the sequence they are using or will use in their Shoebox database. Numbers 1 to 5 are interpreted by the program as from left to right.



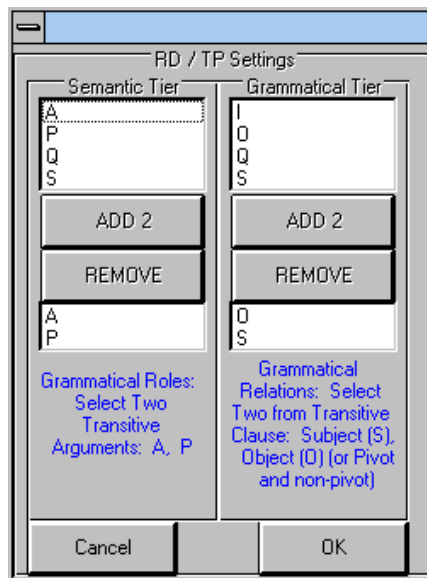
To change the sequencing the user needs to click the mouse on the tier that they want to move up or down in the list. The *UP* and *DOWN* buttons only operate one time. After you have moved a tier up or down in the list you need to select another *tier* or the same *tier* again to move it up or down. Keep doing this until you have the sequence that you want. When you have completed the sequencing, click on the *OK* button to store the sequence settings and return to the previous dialogue box. Click the *Cancel* button if you want to leave the original settings the way they were.

7.2.2 Options

This accesses the *Tracking Options (Colors and Symbols, etc.)*. This option can also be accessed from the **SETTINGS** menu under the options with the same name. This dialogue box is a tabbed dialogue box with three tabs providing subsidiary options. The first tab that opens as the default is labeled *Tracking Options*. This is the tab most likely to be used, and this is explained in more detail below. The second tab is labeled *Participant Settings* and is used to create new participants or open a file with stored participant settings to use in conjunction with the list box on the tool bar when viewing the 2D Map. The third tab is labeled *Screen Settings*. This option is only for users that have their monitor set at a lower resolution than the MDA program was developed on. These tabs are all explained with further detail below.

7.3 RD/TP Quantification Settings

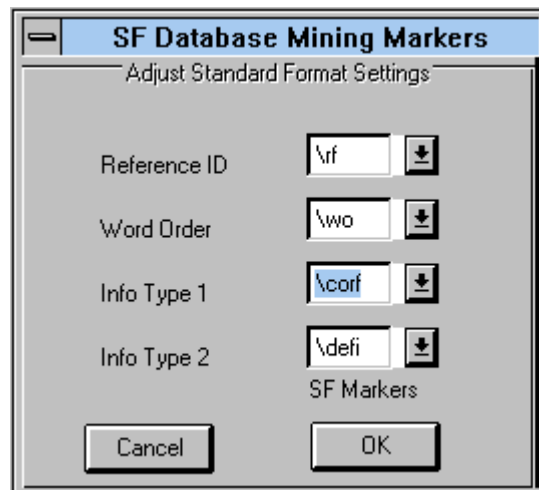
The *RD/TP Quantification Settings* dialogue box is only necessary if the user is using the *Quantification* tool, and specifically is only used in the Topic Persistence and Referential Distance calculations for topic continuity. Two of the *tiers*'s abbreviations' lists used in the *MDA Tag Positions and Codes* dialogue box are copied here into the two upper boxes. Only two abbreviations should be added to the list box below for each tier. These will always be the main two grammatical roles of a transitive clause, and the two main grammatical relations (or pivot and non-pivot). First the user selects the abbreviation in the upper list box and clicks the button labeled *ADD 2*. Then the abbreviation is added to the list box below. In order to remove an abbreviation from the lower list box, first click on the abbreviation that you want to remove and then click the *REMOVE* button.



Click on the *OK* button to store the new or revised settings. Click on the *Cancel* button to ignore any changes or to simply return to the previous program location.

7.4 SF Database Mining Markers

The *SF Database Mining Markers* dialogue box has four combo boxes for the user to set. The first setting is labeled *Reference ID* is the record marker for each clause unit that is tagged (in Shoebox this is the interlinearized unit viewed as one window). If the user has used a different setting than the default `\ref` then the user should change this standard format marker to match his or her database. The next three options are used the same way, but these are optional. These are for marking fields that will be tracked in the span analysis in the 2D Map viewing window. The first of these three which follows *Reference ID* is *Word Order* with the default `\wo`. Again, the user may change the default to match his or her database's standard format field marker. The following two combo boxes are labeled as *Info Type 1* and *Info Type 2*. These are designed for the linguist to track discourse or syntactic information in the 2D Map (more options should become available in the future).



Each of the four combo boxes has two default choices available to use, and one that the user can record if the other two defaults are not used by the user.

7.5 2D Map: Slate Span Group Settings

Span Analyses are set in the dialogue box labeled *Standard Format Settings*, and sub-titled *Adjust Field Codes for Contrasting Spans*. This dialogue box is accessed from the **SETTINGS** menu under the option called **2D Map: Slate Span Group Settings**. This dialogue box has three tabs labeled *Word Order*, *Info Type 1*, *Info Type 2*. Under each of these tabs are two identical boxes to contrast span types for each tab category. For example under word order you may put SVO and SV under *Span Code Group 1*, and put VOS and VS under *Span Code Group 2*. These abbreviations are entered under **/wo** if the user wants to track the spans contrasting these two different word order types. The same principle works for the tabs labeled *Info Type 1* and *Info Type 2*. These are given generic names to reflect the possibility of tracking either syntactic or discourse level information. Each of these tabs also has a pair of boxes that can be used to contrast two groups of information in the spans. Currently the different span groups are marked with different colors, and are labeled below the slate zone to remind the user which spans are which.

Abbreviations are entered in the upper box. Then the user clicks the *ADD/EDIT* button and the abbreviation is added to the lower list box. To remove an abbreviation from the lower list box, click on the abbreviation so that it is highlighted, then click on the *REMOVE* button and the abbreviation will be removed.

Click on the *OK* button to save changes to all of the information under all of the tabs. Click on the *Cancel* button to ignore any changes and/or to return to the previous location in the MDA program.

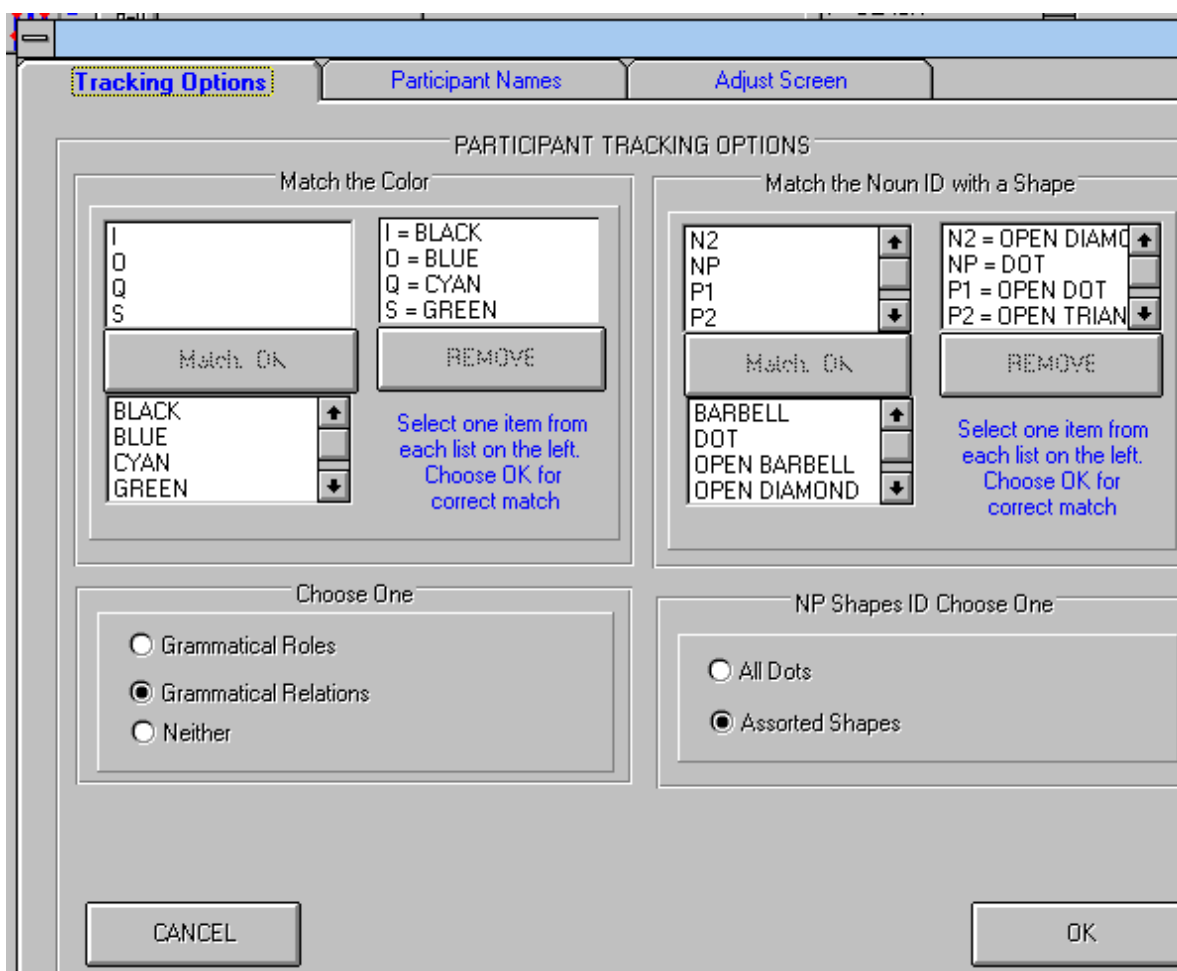
7.6 Tracking Options (Colors and Symbols, etc.)

The user should note that these tabs should be treated as separate dialogue boxes. Before leaving a tab you should click the *OK* button if you want to save changes you have made on that tab.

7.6.1 Tracking Options Tab

The *Tracking Options* tabbed dialogue box can be accessed from **SETTINGS** under the option called **2D Map: Slate Span Group Settings**. This dialogue box can also be accessed under the option button in the *MDA Tag Positions and Codes* dialogue box. The *Tracking Options* tab is one of the most useful and powerful settings available in MDA (and one of the most complicated or detailed of all the settings available in MDA). *Tracking Options* has two major areas that affect colors and shapes on the 2D Map. The colors and shapes are not merely a matter of making the participant tracking look pretty, but signifies options that help the analyst to track and compare different patterns and problems in a discourse context. Colors are used to identify participants' grammatical roles or grammatical relations (i.e. abbreviations used for the semantic tier or abbreviations used for the grammatical tier). Different shapes can be associated with different noun phrase types. By using colors and shapes strategically, the user can see a lot of information coded in the participant tracking area of the 2D Map.

The upper left box under the color section labeled *Match the Color* and the upper left box under the shape section labeled *Match the Noun ID with a Shape* take their information from the abbreviations set in the *MDA Tag Positions and Codes* dialogue box. The area below the *Match the Color* zone titled *Choose One* allows the user to select either grammatical roles (i.e. the semantic tier), or grammatical relations (i.e. the grammatical tier), or neither of the above (in which case a single color is used). Note that it is not possible to view both grammatical relations and grammatical roles at the same time. Selecting one of these options copies the appropriate abbreviations into the upper left box under *Match the Color*. To select a color the user clicks in a color in the list box with the colors. Then the user selects the *Match OK* button. The match is then listed in the upper right list box in the area titled *Match the Color*. To remove a match click on the box to highlight the match to be removed. Then the match is removed. There are eight colors to choose from. One color can be used for more than one association.



Note that if you are printing with a black and white printer that you may get a better printout if you use darker colors or match black with all of the choices (beware that some light colors may not show up on a black and white printer).

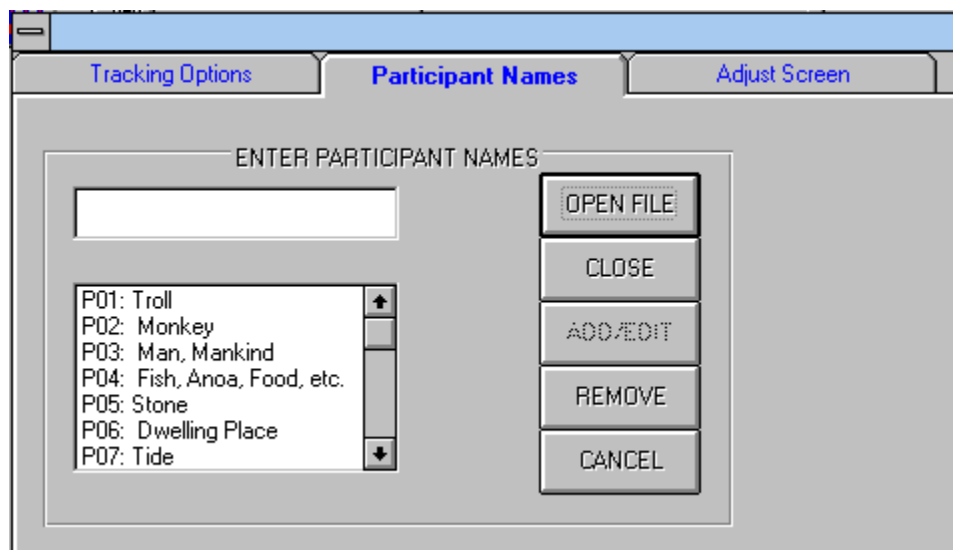
Shapes are associated with Noun Phrase types in the same way as colors are associated with a grammatical role or a grammatical relation. In order to associate shapes make sure the *Assorted Shapes* option is clicked in the area labeled *NP Shapes ID Choose One*. When this option is clicked then the Noun ID abbreviations set in the *MDA Tag Positions and Codes* dialogue box become available. There are currently 10 shapes to choose from, however several shapes are not drawing properly yet, so you may want to stick to the more conservative shapes. To match a shape with a noun type click on each of the lists so that they are both highlighted and then click on the *Match OK* button. Then the match is put in the upper right list box. In order to remove an association click on the association so that it is highlighted and then click on the *REMOVE* button. Shapes can be associated with more than one noun phrase type.

7.6.2 Participant Names Tab

This tab allows the user to store participant names that are used in an analyzed text. This enhances the use of the 2D Map by providing the name of the participant that the mouse is clicked on or last clicked on. Also the text box on the tool bar is a list that the user can scroll up or down on to manually identify which participant is which.

The first time that a text is converted from Shoebox to MDA's database, the participants are numbered and named as 1 Participant, 2 Participant, 3 Participant, etc. The user may edit these names at any time by entering a participant name in the upper text box. Then click the *ADD/EDIT* button to add the participant to the list below. This should be done sequentially to keep the correct order (this is important for the tool bar list box to correctly identify the participant's selected on with the mouse). To remove a name from the list

click on the participant and click on the *REMOVE* button. When you have entered/edited any or all of the participants click on the *CLOSE* button to store and save the names.



8. THE MDA DATABASE (OR *.MLT)

The MDA database will have MDA as the file extension name by convention (i.e. it is not required). The MDA database is a database created from a Shoebox database (or a non-Shoebox database) which has MDA annotations and abbreviations marked in special fields following SIL's standard format marking convention. The MDA database however is simply a comma delimited database, where each line is a different record and commas delimit each field that is recognized by MDA. The MDA database should not be tampered with. Presently there are no special markers identifying the database as an MDA database, so if the user incorrectly selects a file that is not an MDA database the program will try to process the file, and it should fail at some point along the way. The failure should not crash the program but simply provide no results that would otherwise be expected.⁴

9. THE MDB DATABASE (OR *.DB)

I recommend that the user use the extension MDB for the MDA tagged Shoebox database (however it is not required). This will provide the user with a simply way of distinguishing Shoebox databases that have not been annotated with the MDA conventions. Also the file dialogue box will automatically look for the extension of MDB as the default when the user is ready to process an MDB Shoebox database under the Shoebox tool.⁵

9.1 Introduction to MDA Tagging Methodology

It takes a bit of organization and planning to know what will be all the codes to use for each category. You will find that the methodology itself is not difficult to learn. It is the decisions that you will make on an overall framework and clause by clause analysis which is the actual difficulty. I will tell you the five pieces that you need to code. You may want to look at my article (Quick 1996) which diagrams the tagging model. However, now you don't have to strictly follow the codes I presented there (and the Demo version would give you some basic results if you followed that tagging system strictly).

⁴ Note that files with the extension *.mda are reserved extensions for later operating systems such as Win 98 and XP. This is actually not a problem if the user wants to use these extensions. In order to avoid problems the user may want to use another extension name such as *.mlt (which is an abbreviation for *multi-linear text*).

⁵ Note that files with the extension *.mdb are reserved extensions for later operating systems such as Win 98 and XP. This is actually not a problem if the user wants to use these extensions. I suggest if the user wants to avoid this problem to just use the designation *.db as is used with Shoebox.

I will list the five areas that have to be coded into one field for each participant--you can use your own abbreviations--I will suggest some obvious ones, that you can adjust as you wish. The important point is to keep the abbreviations for each of these five areas kept in the same sequence (i.e. 12345, or 54321, etc.), and to always use the same abbreviation for the same thing (i.e. no typos!, etc.). For an initial test I suggest following the sequence I list for as 1 to 5 as left to right. I use the "tier" terminology for labeling purposes here (which you can virtually ignore), but they are concepts from the Multilinear Discourse Theory I have been developing:

1) **Semantic Tier** (also known in Dixon's terms as: A, S, O): Depending on which theory these are called grammatical or semantic roles (but for now I still place it in my theory as the Semantic Tier). I have used A (for the "Agent" of the transitive clause, or the actor), S (for the Subject or single core argument of the intransitive), and P (for the Patient of the transitive clause--or undergoer). For obliques, simply Q.

2) **Topicality Tier**: This is not critical to use in the MDA program yet. This is the fuzziest part of my theory. I have used it for "background" and "foreground" (what I had called Proximate and Obviate in Pendau, or most prominent and second most prominent). If you decide not to use this, use a dummy code to fill this slot (like an asterisk). I am looking for input for better use and development on this.

3) **Reference ID**: What form does the noun phrase take? NP (noun phrase that is not a pronoun or other specification you use), PR (pronoun; you can divide this into different pronoun sets by using different abbreviations.) ZR (I've used for zero anaphora), RC (I've used for relative clauses). Russ Cooper and I came up with an abbreviation slot for verb agreement languages AG (actually I don't remember the exact symbol--but I think this is what you will want to use when the subject is not on the surface in the clause, but its still marked on the verb, since its not really zero?!). This last one is really a tagging issue--so you may be able to help me solve this issue and find the best solution (or possible choices).

4) **Pragmatic Tier**: Is the NP definite or indefinite? I just have used a minus sign (-) or a plus (+) sign, Wayan has been using Indef or Def in coding his Balinese texts. Alternatively you optionally can use this for givenness or specificity, etc.

5) **Grammatical Tier**: What are the grammatical relations for this participant? Usually I use I for Intransitive Subject, S for grammatical subject of a transitive clause, and O for Object of a transitive clause. I've also usually indicated obliques redundantly here as Q. For an indirect object, maybe IO.

It is best to keep the sequence of these five "zones" coded succinctly, but avoid ambiguity. If you use Q two times in two different zones, MDA Beta version can disambiguate, but obviously you don't want to do this too many times as the code won't mean anything to you if you do. You can look at my CARLA paper for samples of these five zones that I've used for Pendau. You can probably follow fairly closely the same tagging system I developed for Pendau, but just add a few extras. You may want to lengthen each abbreviation to two or three characters--that doesn't matter now. I repeat, its not restricted to only seven character positions as was the MDA Demo B+ version.

These five zones are coded for each participant of each clause--normally only 2 to 3 participants (and or props) will be coded for each clause. In the MDA Beta version you no longer need to put in a field for each participant in every record. Here's a short example:

```
\ref Text 001(required record marker)
\txt Vernacular text goes here (but is not necessary to run the program--this is only if you're using a
      Shoebox database)
\mr morpheme break line used in Shoebox for example (Optional--not necessary for MDA)
\ge English gloss (Optional--not necessary for MDA)
\ps Part of speech (Optional--not necessary for MDA)
\fr free translation (Optional--not necessary)
\wo AVO (Word Order of clause--optional in MDA but usually for Span Analyses)
\it1 DEF (Discourse information type 1--contrast for example definite and indefinite in Span Analyses--
optional in MDA)
```


- \it2 SS (Discourse information type 2--contrast for example Same Subject versus Different Subject in Span Analyses -- Optional in MDA)
- \p2 12345 (necessary for each participant in the clause--numbers represent the five zones that you use for the abbreviations)
- \p5 12345 (necessary for each participant in the clause--numbers represent the five zones that you use for the abbreviations)

Notice that many of the database fields are optional--what are strictly required are the participants for each clause. Only list the participants that are relevant in any one clause. Although this may be controversial, in following Givón's methodology, I think you only should code the core arguments, and oblique arguments which are the main references of its constituent. For example in "my dog" I would code the dog as the participant and ignore "my" as irrelevant (in fact I believe its a mistake to do so--although it may have some significance, I don't believe it fulfills the topic continuity methodology (I could be wrong--I would glad to be enlightened or hear another opinion).

Also notice, that one main enhancement in the MDA Beta version is the availability of three optional fields for discourse information (actually three types above not just two--most analysts will want word order I believe). In the discourse information fields you will be contrasting two pairs of information. Since Pendau has SVO and VOS I code all the possibilities in this one field. Then in a user setting in MDA Beta version I list them by grouping them into two different categories. Then MDA shows you graphically by colored spans next to the participant tracking what the spans are! It beats doing this by hand! However, it doesn't provide statistics on anything yet. You are welcome to begin giving me input on other kinds of statistical information you think would be helpful to have automatically calculated by the MDA program.

9.2 Preparing the Shoebox database for MDA Analysis

One technique that is recommended to the discourse analyst is to group the coding abbreviations that will occur for each field marker, and to group the coding abbreviations that the user will use for each tier. There are basically two categories of standard format markers (excluding the record marker itself which is also required). The first group of standard format markers are used for span analyses. The standard format markers used are up to the user, but they have to be set in the settings dialogue box called *SF Database Mining Markers* (the record marker is also set in this dialogue box). Information of almost any sort can be tracked by these markers. One marker that most analysts will want to track is word order. The different word orders are then abbreviated in this field. The abbreviations are then recorded in the MDA settings option called *2D Map: Slate Span Group Settings*, and are placed in binary oppositions (I hope to allow for more than binary oppositions in the future). Another field marker could be used to identify non-event and event clauses. Yet another field marker could track other discourse information that the user wants to track or trace in context.

The second group of standard format markers uses the convention of *\pI*, where the *p* symbolizes *participant* of a clause and the number can vary from 1 to 999. The *p* is required along with the number following it in order to identify this is a field where the five tiers are encoded as abbreviations. Every clause will have a participant that needs to be tagged with information from each of the five tiers. How this is abbreviated is up to the analyst, but the abbreviations must be consistent, and they must always have the tiers in the same sequence (the sequence can be set in the *Tier Tag Sequence* option which is accessed from the *MDA Tag Positions and Codes* dialogue box).

Participants that occur more than once in the clause can both be tagged. To do this the field marker is repeated (i.e. *\pI* is reproduced on the following line) and then each instance is coded differently according to the different information it represents. Another improvement over the Demo B+ version is that now it is only necessary to record fields for those participants that occur in the clause (it is no longer necessary to have empty fields for all other participants).

As I explained in the CARLA paper (Quick 1996), if the user is using Shoebox then the abbreviation entry process can be automated by using the range parameters for each field. This allows the user to stay consistent and not to enter wrong annotations.

The user should review some sample MDB databases that are provided to help understand how the annotated tagging is performed if there are still any questions of doubt.

10. GETTING ACQUAINTED WITH MDA

For the first time user, I am recommending that he or she run MDA and practice using the MDB and MDA files that come with the installation (first time users can try the first tutorial lesson presented earlier in this manual). By practising with these sample database files you will begin to become acquainted with the capabilities of MDA, and it you will show you that the program is running properly on your computer. I strongly advise that you make sure that the sample files work before you begin any text tagging.

Also check out all of the settings to see where things are, this will speed up the learning curve for you when you are ready to begin entering your own custom-made settings.

10.1 A Quick and Simple Tagging Approach in Learning to Use MDA

One shortcut that some first time users may want to try is tagging only the NP type. In the participant field it is possible to enter dummy symbols for all but the NP type and still be able to use some of MDA's features. Note that this method will not allow the user to use any of the topic continuity quantification techniques. However the Text Tokens and the 2D Map viewing will still function and give the user a good idea of how the program works without a lot of work analyzing every clause. The lines below demonstrate how this could be approached.

```
\p1 12_NOUN_45  
\p7 12_PRONOUN_45  
\p33 12_ZERO_45
```

The numbers 1,2,4,5 would be used for every single participant entry and entered in the MDA Tag Positions and Codes dialogue box. That is only enter the numeral 1 as a dummy symbol for the semantic tier, the numeral 2 in the topicality tier, the numeral 4 in the pragmatic tier, and the numeral 5 in the grammatical tier. Any symbol can be substituted for these numerals (such as an asterisk, a hyphen, a plus symbol, etc.).

11. MDA TAGGING DIFFICULTIES

There are a number of problematic issues that the linguist will face when attempting to analyze many clauses. Some of these are theoretical issues that are still beyond the scope of the MDA program. The linguist will have to wrestle with how he or she wants to represent problematic clauses, just as he or she would in describing them in a grammar. One major problem that cannot be solved in a 2D representation of clauses in context is the issue of "splitting" and "joining" (terms suggested by Russ Cooper (personal communication)). The user will need to make a consistent judgement in coding plural pronouns and plural and collective nouns. The question is in regards to tracing them in a two-dimensional model when they are referred to separately and when they are referred to jointly. What is the best technique or method of coding and tracing them in a two dimensional participant tracking model (or "map" in the MDA program). My suggestion is to always keep separate as many distinct participants as are referred to uniquely, and whenever they are coded as a plural pronoun, then each participant is coded identically, but with its own unique field marker. This however brings a problem with the statistical counting. For now the MDA user must be aware of the limitations of the program, and understand that this problem may skew the statistics. My partner, Kevin Madden and I, hope to work on a three dimensional model for MDA version 2, sometime within the year following the release of version 1. We will then begin to attempt to resolve this problem among others.

12. UNDERSTANDING THE TOPIC CONTINUITY STATISTICS PRODUCED IN MDA

The statistical suite available in MDA currently produces 7 different topic continuity analyses of texts [see supplemental notes in the following section for additions and current status of statistical options available]. The data is produced currently in a rough text format. This is available to cut and paste directly from the MDA viewer, or from the text file that is created. Currently the user may need to format the data in

a table in a word processor in order to make better use of the data. There are examples in the appendices of how the data can be displayed in a word processing table or chart, and using statistical graphing packages available in programs. I have successfully used the graphing capabilities available in MS Word. Since the data is displayed in a rough table in basic text format, if the analysis is fairly complex, the user will need to carefully line up the data with its correct column.

13. SUPPLEMENTAL NOTES TO THE MDA BETA USER'S GUIDE: AN UPDATE

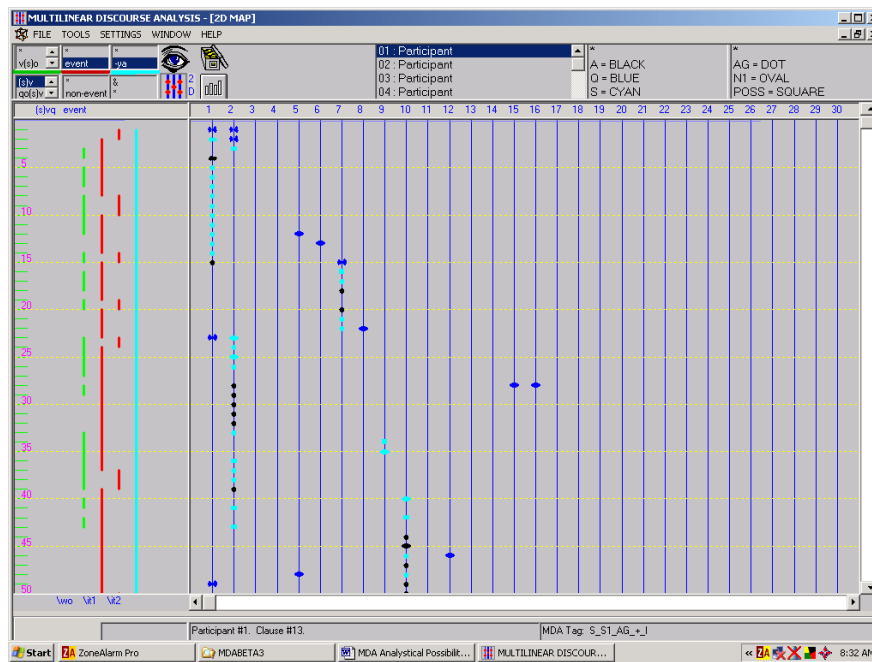
13.1 MDA Analytical Possibilities

This section gives a short list of some common problems that the MDA program can be applied towards solving.

- 1) Verify and/or identify the transitivity of a second transitive construction through topic continuity statistics (use the Givón statistical approach)
- 2) Identify the selection criteria between two distinct transitive construction types (use the Dryer statistical approach)
- 3) Identify the reason for pragmatic word order differences such as between SVO and VOS (see Quick 2006 for VP fronting as a possible explanation for the Pendau language, and which may be applicable to other similar structured languages—also compare Levinsohn and Dooley articles)
- 4) Verify a transitive construction is not an antipassive (necessary to disqualify an ergative analysis in some cases) [use the Givón statistical approach]
- 5) Levinsohn's S1-S4 statistics (newly proposed for MDA) [tag these as one of four abbreviations in the 'topic' tier, i.e. S1, S2, S3, or S4] It is not required to tag these unless the user wants to do this analysis. If the user does not want to tag this tier, code as an asterisk or some other unique abbreviation.
- 6) Examine the raw profile of the frequency of particular participants according to their NP type, according to their occurrence as a semantic role/macro role, or according to their occurrence as a grammatical relation.
- 7) There are many other analyses possible by using the interactive capabilities and making various 'what if' hypotheses to test these. One of the benefits of the MDA program is to help the user make rapid shifts and/or refine their hypothesis testing or analysis, and to analyze a higher number of texts more easily.
- 8) An additional goal for the MDA tags in the shoebox/toolbox database can be to search for specific examples of text data with a specific code or coding combination. This can be used for various grammatical analyses and for finding examples to use in your research paper or other research goals.

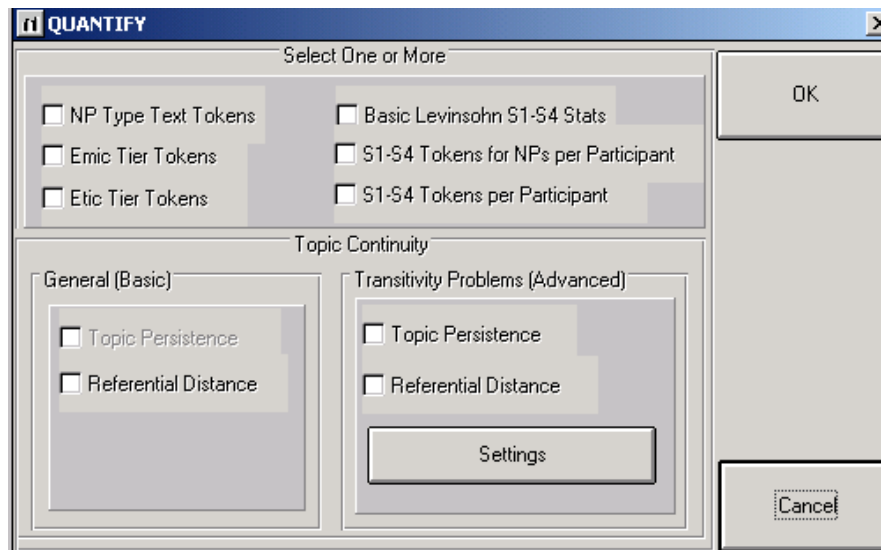
13.2 New Functionality in MDA 1 (beta 4.1—2005-2006)

There are six new list boxes near the top left corner to use in the 2D Map interactive viewing. Each pair of boxes dynamically tracks each pair of contrasts the user is tracing in the span analyses. The top box of each pair is the left trace line of a particular color trace, and the lower box of each pair is the right trace line of a particular color. In between each box is the same color used for each span analysis trace. When the mouse is clicked on a particular clause the item is highlighted in either the top or lower box of the pair to help the user quickly interact and identify what is being traced (note: there may be still some instances when there are mismatches or highlighting is not turned on or off...currently if the user enters a different number of settings then have been matched, then the user may observe some mismatch occurring). If there is an extra space at the top of the participant list, then open up the participant tab in the tracking options and delete the empty space at the top of the list, and then the match should match correctly.



13.3 Six new quantitative features with a total of 9 quantitative features

Each statistical option is summarized below, following the screen capture showing the choices now available in MDA Beta 4.1 (note that the 10th feature, a basic topic continuity analysis for topic persistence is still on the to do list).



13.3.1 GENERAL STATISTICS

NP Type Text Tokens—Click on this to get a statistical summary of how many times a NP type occurs for each participant. This is a simple matrix that can be copied and pasted into a word processor for further formatting.

Emic Tier Tokens (=Grammatical Relations)—Click on this to get a statistical summary of all the occurrences of each grammatical relation or other NP argument coded by the user by their respective types.

Etic Tier Tokens (=Semantic Roles/Grammatical Roles/Macro roles)—Click on this to get a statistical summary of all the occurrences each semantic role/macro role or other NP argument coded by the user by their respective types.

13.3.2 LEVINSOHN S1-S4 STATISTICS

Basic Levinsohn S1-S4 Stats—Click on this to get a matrix statistical summary of S1-S4 occurrences for each NP type [These are coded in the topicality tier as one of these: S1, S2, S3, S4]

S1-S4 Tokens for NPs per Participant—This is similar to the Basic Levinsohn approach, but examines the statistics according to each specific participant and according to each NP type.

S1-S4 Tokens per Participant—This is similar to the Basic Levinsohn approach, but examines the statistics according to each specific participant.

13.3.3 TOPIC CONTINUITY OPTIONS

General (Basic)

Topic Persistence—[code not yet written—see advanced topic continuity options to conduct some topic persistence analyses]

Referential Distance—This will give a detailed summary matrix for every referential distance possibility 1) based on the emic tier, 2) based on every NP type, and 3) based on the etic tier. This allows the user to focus on a particular referential distance, or every possibility. [See Transitivity Problems (Advanced) for a more sophisticated use of some topic continuity techniques] This technique follows Givon's basic techniques described in his 1994 book.

Transitivity Problems (Advanced)

Topic Persistence—Click on this to compare two transitive constructions (or one transitive construction and one possible transitive construction) in order to get a matrix with topic persistence statistics which compare both constructions. This technique follows Givon's basic techniques described in his 1994 book. Also see appendix 6 in the MDA User's Guide for more discussion, *Generic Description of Topic Continuity for Discourse Analysis*. In addition to the basic matrices produced following Givon's approach, this analysis also produces Dryer matrices. These matrices compare the two main arguments of a transitive (or a putative transitive construction) simultaneously.

Referential Distance—Click on this to compare two transitive constructions (or one transitive construction and one possible transitive construction) in order to get a matrix with referential distance statistics which compare both constructions. This technique follows Givon's basic techniques described in his 1994 book. Also see appendix 6 in the MDA User's Guide for more discussion, *Generic Description of Topic Continuity for Discourse Analysis*. In addition to the basic matrices produced following Givon's approach, this analysis also produces Dryer matrices. These matrices compare the two main arguments of a transitive (or a putative transitive construction) simultaneously.

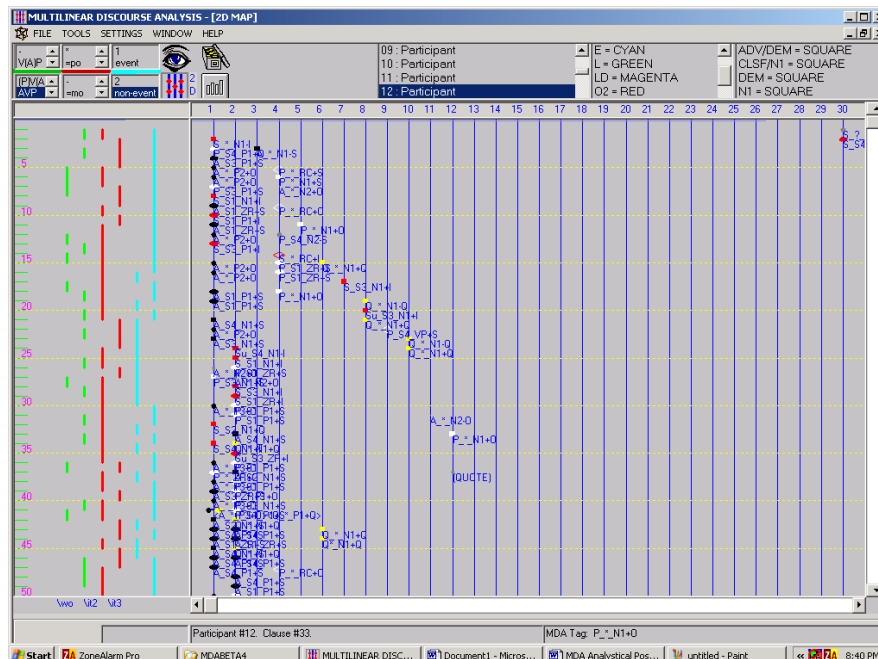
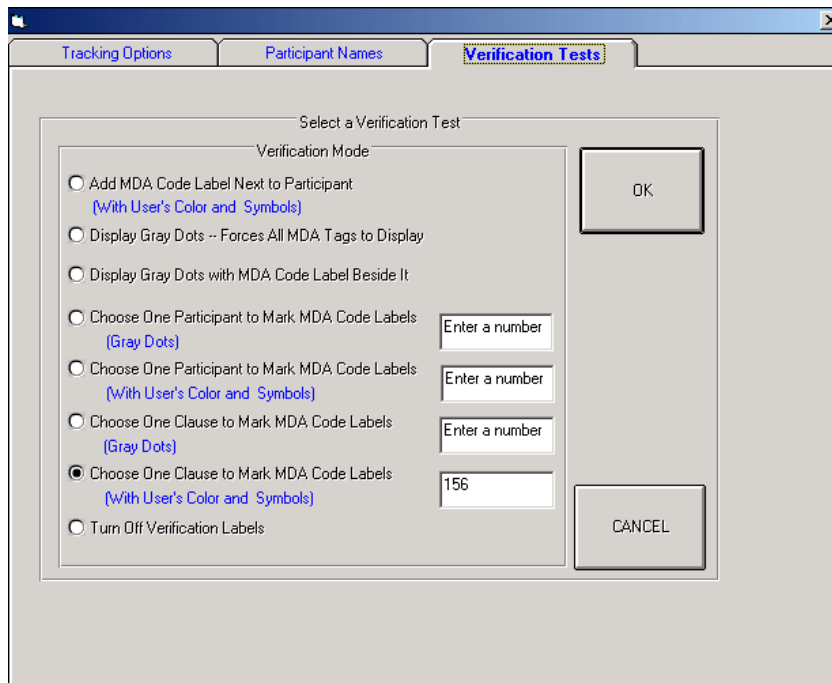
Other New Features:

Verification tests provides a suite of eight features to confirm or verify whether or not the MDA tag is being displayed properly or even at all. Display problems of symbols includes: forgetting to enter an abbreviation in the MDA settings. Another new verification test that works automatically is the display of gray dots in the normal 2D Map display when there is a failure of some sort to display it according to a color and shape. The gray dots are also options here for the user with a mix of several possibilities. There is still a problem in the program when switching back and forth between choices—sometimes there will be an overlay of the previous color or choices. The work-around is to move up or down at least one increment for a complete 'true' resetting of the current choice.

The screen capture below shows a display of the verification choices for inserting the MDA tag labels beside each symbol. The user will notice that in some texts such as this one, that there is far too much overlap to make sense out of some labels. In this case select a single participant or a single clause for labels to be displayed on. In other texts the participants will be spread out so that the maximum use of MDA tags displayed will be helpful.

There are several reasons that the user may want to use the verification options: 1) The user forgot to enter an abbreviation in one or more of the setting list boxes (these must be the same abbreviations used in the MDA

annotations in the participant fields), 2) the MDA program has failed to disambiguate ambiguous abbreviations (the user should make sure that very few abbreviations are the same or similar) in each of the different tiers), 3) the MDA program may mistakenly analyze an abbreviation (this could possibly happen if there is also an ambiguous set of abbreviations), 4) the user has mistyped an abbreviation in the original database or improperly coded a standard format field.



Features removed or disabled from previous beta versions:

- Screen resolution tab has been replaced with verification features
- It is now unnecessary to designate the location of your html browser. Clicking the online help will open the MDAGUIDE.htm in your default browser automatically. [Note that the current version of the online help is outdated, although it may still be of some use. It is still in a very rudimentary form.]

New Features in process, but not available yet:

- Overlay trace will implement a third information type (it3) that functions the same way as the span analyses (in the 'slate' area), however it will display on the 2DMAP as a color for each clause.
- Basic topic continuity for topic persistence
- Expanded printing options (currently limited to one choice)

References for using MDA's Topic Continuity Statistics

- Dryer, Matthew. 1994. The discourse function of the Kutenai inverse. In Talmy Givón, ed. *Voice and Inversion*. Amsterdam: John Benjamins Publishing Company.
- Gildea, Spike. 1994. Semantic and pragmatic inverse: "Inverse Alignment" and "Inverse Voice" in Carib of Surinam. In Talmy Givón, ed. *Voice and Inversion*. Amsterdam: John Benjamins Publishing Company.
- Givón, Talmy. 1994. The pragmatics of de-transitive voice: Functional and Typological Aspects of Inversion. In Talmy Givón, ed. *Voice and Inversion*. Amsterdam: John Benjamins Publishing Company.

References for Analyzing Levinsohn's S1-S4 Statistical Approach

- Dooley, Robert A. and Stephen H. Levinsohn. 2001. *Analyzing discourse: a manual of basic concepts*. Dallas: Summer Institute of Linguistics
- Levinsohn, Stephen H. 2003. Analysis of narrative texts. Manuscript, SIL International.

Bibliography

- Givón, Talmy, ed. 1983. *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins Publishing Company.
- _____, ed. 1994. *Voice and Inversion*. Amsterdam: John Benjamins.
- Grimes, Joseph E. 1975. *The thread of discourse*. The Hague: Mouton.
- Larson, Mildred. 1984. *Meaning-based translation: A guide to cross-language equivalence*. Lanham, Md: University Press of America.
- Longacre, Robert E. 1983. *The grammar of discourse*. New York: Plenum.
- Longacre, Robert. 1989. Two hypotheses regarding text generation and analysis. *Discourse Processes* 12:413-460.
- Quick, Philip. 1996. Multilinear discourse analysis software demonstration. In H. Andrew Black, Alan Buseman, David Payne, Gary F. Simons (eds.), *Proceedings of the 1996 General CARLA Conference, November 14-15, 1996*, pp. 291-309. Waxhaw, NC/Dallas: JAARS and Summer Institute of Linguistics.
- _____. 1997. Active and inverse voice selection criteria in Pendau, a Western Austronesian language. In *Proceedings of the Seventh International Conference on Austronesian Linguistics*, pp. 461-482. Amsterdam/Atlanta: Editions Rodopi B.V.
- _____. 2002. A sketch of the primary transitive verbs in Pendau. In *The historical and typological development of Western Austronesian voice systems*, ed. by Fay Wouk and Malcolm Ross. Canberra: Pacific Linguistics.
- _____. 2003. A grammar of the Pendau language. PhD dissertation. Canberra: The Australian National University.
- _____. 2005. Topic continuity, voice, and word order in Pendau. In *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*, I Wayan Arka and Malcolm Ross, eds., pp. 221-242. Canberra: Pacific Linguistics.
- _____. 2006. Is there a VP in Pendau? Paper presented at 10-ICAL, January 2006, Palawan, Philippines.
- _____. In prep. *A discourse-text profile of Pendau*. drafted manuscript.

Basic Glossary

MAP A 2D graphic representation of a text. This is the right side of the screen where the participant tracking is displayed.

SLATE This is the area of the 2D Map which displays the span analysis. The span analysis contrasts 3 pairs of contrasts in parallel to the participant tracking on the right hand side of the screen.

PARTICIPANT A participant is any referent in a clause. I use the term participant in very broad terms to include participants and/or props that are generally designated in the clause as a core argument or an oblique argument. Participant fields in an MDA tagged Shoebox database will begin with the letter *p* and be followed by a number, such as: \p1, \p5, \p9, \p33. Participants are used in participant tracking in the 2D Map. Participants are displayed with various colors and symbols. 30 participants can be tracked in one screen view. A vertical blue line below a numeral at the top of the 2D Map designates the continuity of an individual participant (i.e. a participant or participants that are tagged that way in the MDA database).

User's Notes: