HyperLAT, the Lexicon Acquisition Tool of the Project DBR-MAT

- Introduction and Manual -

Walther v.Hahn
Computer Science Department. University of Hamburg. Germany

Abstract

HyperLAT (HyperText Lexical Acquisition Tool) is a software tool for producing syntactically correct lexical structures within the definitions of the project DBR-MAT. The most important features of the system are (1) multilinguality, which means definition of lexical correspondencies and co-occurrences, and (2) definition facilities for linguistic features, which allow for linguistic (re)definition of (new) languages. Because of the script language HyperText, the system is running on Macintosh computers only.

Zusammenfassung

HyperLAT (HyperText Lexical Acquisition Tool) ist ein computerlinguistisches Werkzeug zum Erstellen syntaktisch korrekter Lexikoneinträge in den Definitionen des Projekts DBR-MAT. Wesentliche Merkmale des Systems sind seine Multilingualität (Angabe von lecikalischen Entsprechungen und Kookkurenz) sowie die freie (Re-)Definierbarkeit von linguistischen Merkmalen einer (neuen) Sprache. Das System läuft wegen der benutzten Skriptsprache HyperText nur auf Macintosh Computern.

Table of Contents

1 Overview ... 3
1.1 Functionality ... 3
1.2 Components ... 3
1.3 General Functional Description ... 3
2 Read this important information first! ... 5
3 A General Introduction to HyperLAT ... 6
3.1 General remarks ... 6
3.2 The HyperLAT Acquisition Card ... 7
3.3 The Grammar Maintenance Cards ... 7
3.4 The Save Grammar Data Cards ... 7
3.5 The Load Grammar Data Cards ... 7
3.6 The Lexicon Cards ... 7
3.7 The Translation Card ... 7
3.8 Other Cards ... 8
3.9 Other Useful technical details ... 8
4 Practical Use of HyperLAT ... 9
4.1 How to Use HyperLAT ... 9
4.2 Steps to enter new lexicon entries ... 10
5 The Acquisition Card ... 12
5.1 General remarks ... 12
5.2 Functionality ... 12

6 Translation 14
7 Word Lists 15
7.1 Where is the word list used? 15
7.2 Where does the word list come from? 16
8 Internal Lexicon Card 17
8.1 The lexicon card contains in 7 fields: 18
8.2 The HyperLAT lexicon card and the DBR-MAT lexicon 18
8.3 Incomplete entries 18
8.4 Other fields and buttons 18
8.5 The Option button 19
9 Grammar Maintenance 19
9.1 General information 19
9.2 How to edit grammar rules: 20
9.3 Trubleshooting 21
10 Saving Grammar Data 22
11 Loading Grammatical Data 23
12 Paths and Preferences 24
12.1 System Path 24
12.2 Language Pairs 24
12.3 Top numbers 25
13 Troubleshooting 25
14 Technical Information 26
14.1 General Information 26
15 The HyperLAT Folder 26
16 Info on the DBR-MAT System 27
17 Lexicon Entries in a BNF-similar style (Version 3) 28
18 Examples 30
18.1 Main Entry of a Term: 30
18.2 Main Entry of a Word: 30
18.3 Entry for short reference: 31
18.4 Entry for Inflected Forms: 31
18.5 Entry for Morphological Rules: 32
18.6 Entry for Structural Rules: 32
18.7 Entry for Explanations of Words: 32
18.8 Entry for Translations: 33
18.9 Entry for Syntax: 33
18.10 Entry for Semantic Descriptions: 34
18.11 Entry for Phrases: 34
19 Indices 35
19.1 Highest Lexical Entry-Id: entry_index_g(341). 35
19.2 Highest Morpho Entry-Id: morpho_index_g(74). 35
19.3 Highest Structural Entry-Id: struct_index_g(7). 35
20 File Hierarchy of the HyperLAT Files 35

Table of Figures

Figure 1: Components of HyperLAT 3
Figure 2: The acquisition card 4
Figure 3: HyperLAT Help top card 5
Figure 4: Schematic overview 6
Figure 5: Acquisition card with Romanian setting 9
Figure 6: Translation card 14
Figure 7: The word list card 16
Figure 8: The lexicon card 17
Figure 9: Grammar maintenance card 20
Figure 10: Saving grammar data card 22
Figure 11: Loading grammar data card 23
Figure 12: Paths and preferences card 24

1 Overview

1.1 Functionality

HyperLAT (HyperText Lexical Acquisition Tool) is a tool to produce syntactically correct lexical structures. The most important functional requirement in the corresponding project DBR-MAT was its flexibility and multilingual functions.

1.2 Components

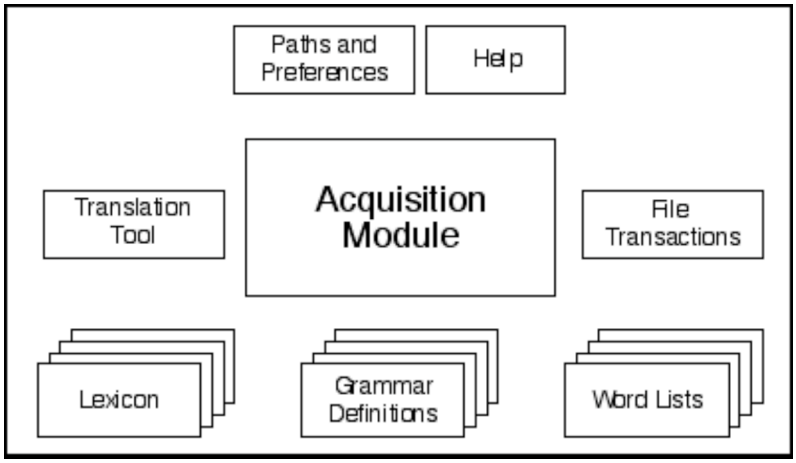

Figure 1: Components of HyperLAT

1.3 General Functional Description

HyperLAT allows for a flexible acquisition of multilingual Words and terms for DBR-MAT, including ist syntactically correct storage..

On the central "HyperLAT Acquisition Card" the user is supposed to enter new lexical entries with all their obligatory and additional DBR-MAT relevant features

First, the user has to set the language pair when other languages (and other character sets) have been under work. A field on top of the card indicates the source language.The target language (important only if to enter translations) can be seen from the title of the "Enter Translation" button in the lower right corner. Clicking the button Change Languages will open the "Paths and Preferences" card to define languages consistently.

The user types in either a word or a term, both, simplex or compound, simple or complex.

The button "**Already there**?" allows for checking whether a word is already in the lexicon. If the word is really new, the button has "no!" result. Otherwise the entry is shown on the word list card.

Successively he/she can define the word ***linguistically*** by entering

- the part of speech,
- the gender, if applicable,
- a grammatical/inflectional class (for single words/terms

or

- a syntactic class (for complex entries, i.e. entries with more than one separated element, respectively).

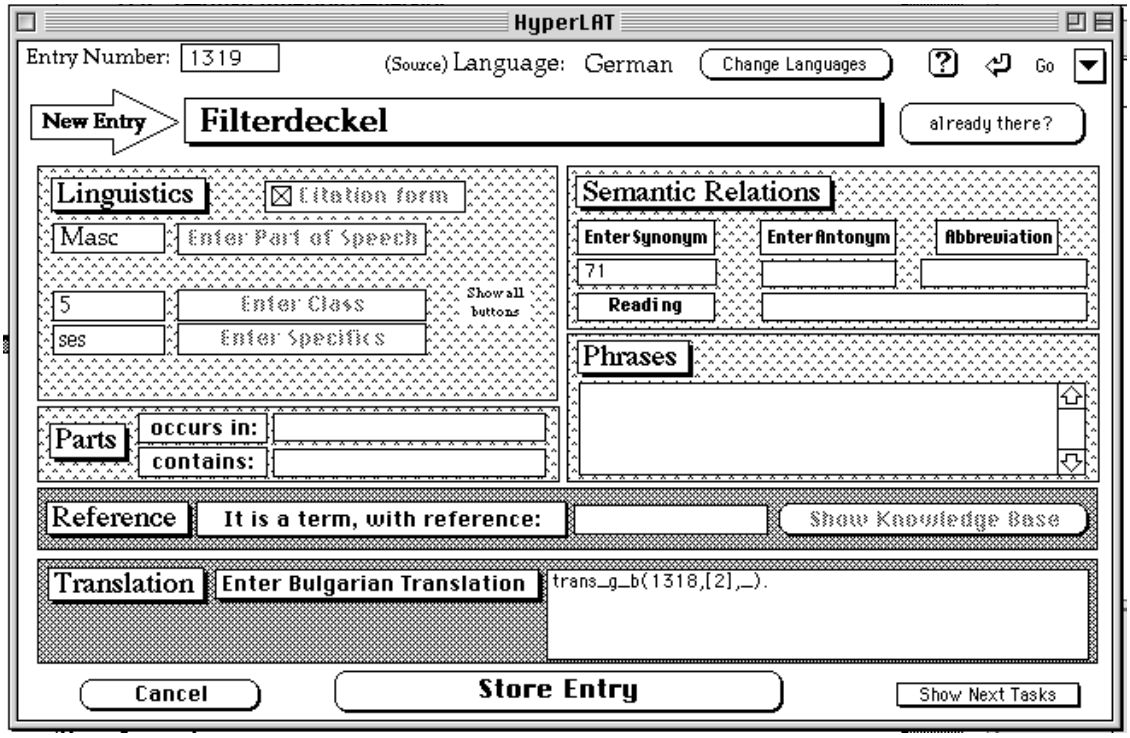

Figure 2: The acquisition card

Only applicable buttons are enabled by the system. For each entry step a specification field (offering the possible choices) is provided, to ensure consistent and homogeneous entries. The necessary fields and buttons are enabled, disabled, shown or hidden according to the grammar of the corresponding language.

Under ***semantics*** synonyms, antonyms and abbreviations can be entered. The requested reference to other words is provided by selection from a "Word List", a summary of all entries and their identification numbers. Additionally you can enter one word to sketch the reading (meaning) of a word for disambiguation. The reason, why: In the word list you may find the line "bank 789" but you do not know whether it is the financial bank or the park bank. The entry "reading" is only used on the word list

For compounds or complex entries you can mark which parts the lexeme **contains** and in which lexemes the entry **occurs**. Again the choice is made from the "Word List".

If you are sent to the "Word List" card and you discover that a word is not yet in the Word List (hence, the lexicon), the opportunity is given to enter this word immediately, because synonyms and antonyms, if not necessary in the lexicon otherwise, can be stored as short entries without detailed linguistic information. After this interruption the user can continue entering the original lexeme.

Reference provides two alternatives (shown mutually excluding): If the button shows "It is a term…", you can enter (in future versions) the corresponding graph id (not yet implemented). If the button is"It is a word…" (it is an everyday word), you can enter a free text explanation in a text field.

The header field in ***Enter ...Translation*** changes its title according to the specified target language (a simple way to see, what is defined). In the text field you will see the formal entry after the definition on the translation card.

A Grammar Definition Card allows for specifying the contents of the fields to click values from. By this technique the system can be changed rather quickly and under work.

Loading and saving facilities produce/load external lexicon files to be used in the DBR-MAT system.

The HyperLAT Manual

Sections 2 - 20 is the complete text of the HyperLAT Help, a separate HyperCard stack with the following top card:

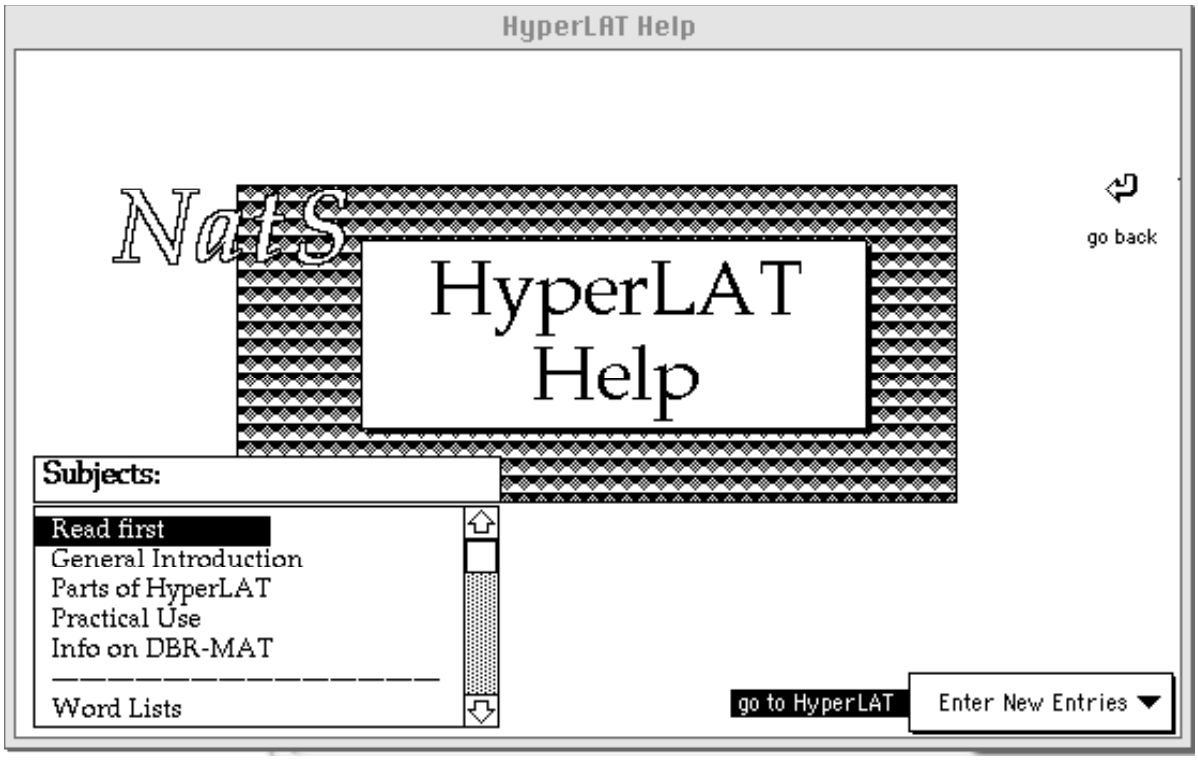

Figure 3: HyperLAT Help top card

2 "Read this important information first!"

The system has certain requirements to work properly:

1. Before starting your work make sure that the paths are set correctly and the languages are defined as needed for the entries. You can check the paths and the languages on the card "Paths and Preferences"

2. All folders in which HyperLat is contained must consist of single names (underscores allowed) (no "HyperLAT System" or something like that).. The name of HyperCard must not be changed

3. The HyperCard system size must be set to at least 2000 K (How? Select the symbol of the closed system, i.e. while the system is not active and press + I . In the opening window you can set the space on the last line)

3 A General Introduction to HyperLAT

3.1 General remarks

HyperLAT is a tool to produce syntacticly correct lexicon entries for DBR-MAT in four languages and their pairs.

HyperLAT and this help file is implemented with the hypertext system HyperCard. The scripts behind buttons or cards are written in the script language HyperTalk. If you are not acquainted with HyperCard, please read the first chapters of the HyperCard manual first or go to the HyperCard Help file (click here if you want to see it). The HyperTalk manual is again a HyperCard document.

To get a very general idea of HyperLAT you can look at the schematic overview. Click the boxes there to go to the info cards. (NB: in HyperCard you cannot move the cursor by using the arrow keys!). The "house"-symbol on all cards in the right upper corner of each card leads - when clicked - to the title card, from which you can reach all important processing cards.

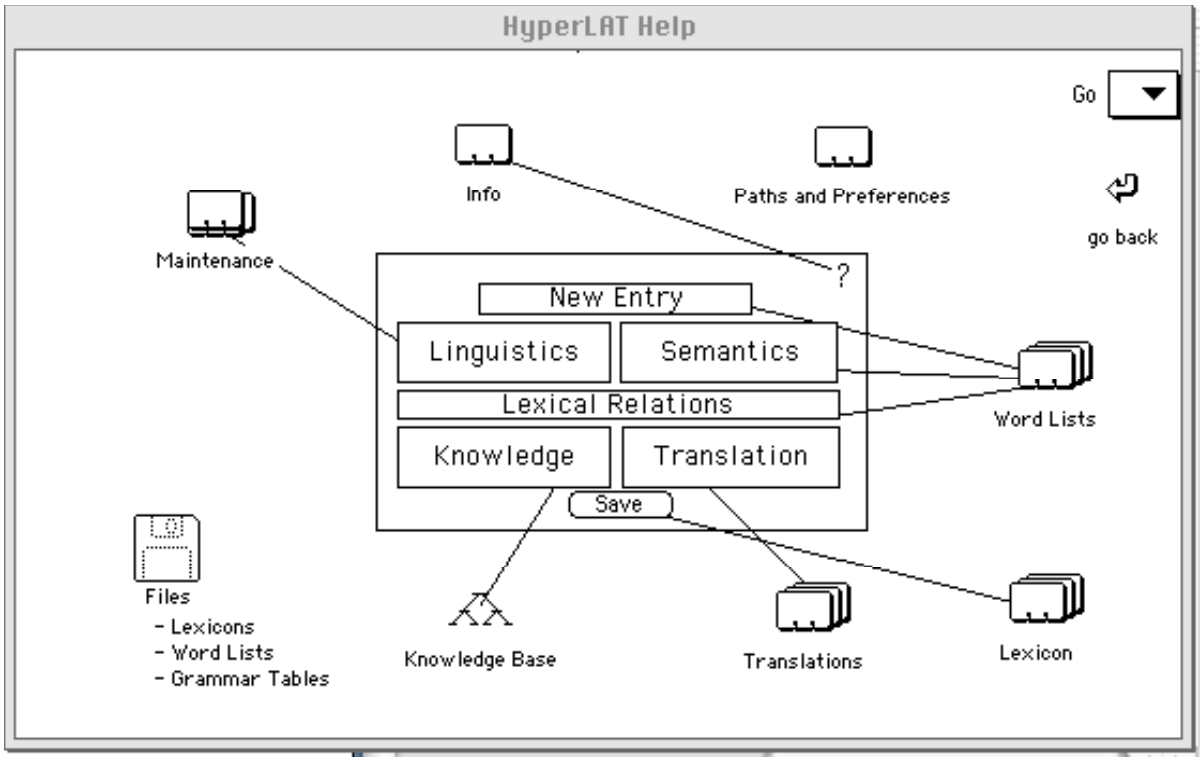

Figure 4: Schematic overview

DBR-MAT is a Machine Aided Translation system programmed in Prolog. To day the system can handle three languages: German, Romanian and Bulgarian.

Before working with HyperLAT you have

(1) to set the path names of the lexicon and the word lists. The system must know, where the new entries should be stored and from where, if existing, word list should be copied.
(2) to choose the language pair for the lexicon entries. The "source language" is the language of the lexicon item itself and the "target language" that of its translations.

For these two processes please read the description of the help card "Paths and Preferences" very carefully (select it from the button "Go" on the top of this card).

In the following paragraphs you will find only short information about the various cards of HyperLAT. For futher details select the corresponding line of the "Go" button on the top of this card (see chapter 5 ff).

3.2 The HyperLAT Acquisition Card

On the Acquisition Card card the user is supposed to enter new entries of the DB-MAT lexicon with all its obligatory and optional features.
Please keep in mind, that a specific language pair is always set and you have to change it when other languages (probably with other character sets) are concerned.

For further information select "Acquisition" from the menu "Go" (see chapter 5)

3.3 The Grammar Maintenance Cards

allow you to write or edit grammar rules (parts of speech, inflection, specifics, and noun phrase syntax) for each language. These tables are used on the Acquisition Card to choose from. The rules are loaded to fields which appear automatically, when a form field must be filled.

For futher details select the corresponding line of the "Go" button of this card (see chapter 9)

3.4 The Save Grammar Data Cards

allow for saving the contents of the grammar fields. They are not saved with the lexicon because they are not parts ot the lexicon but are only used in HyperLAT.

For futher details select the corresponding line of the "Go" button of this card (see chapter 10).

3.5 The Load Grammar Data Cards

load grammar data from external files this may be necessary when you switch from one HyperLAT version to another and want to keep the grammatical definitions.

For futher details select the corresponding line of the "Go" button of this card (see chapter 11).

3.6 The Lexicon Cards

serve as an overview of the entries, which have been produced with the help of HyperLAT so far. New entries can be saved to the "original" external lexicon file later on from there.

For futher details select the corresponding line of the "Go" button of this card (see chapter 8).

3.7 The Translation Card

On the Translation Card you can establish correspondencies between one (or more) words of the source language and one (or more) words of the target language. The definition of these languages on the card "Paths and Preferences" loads word lists of both languages to this card to select from them.

For futher details select the corresponding line of the "Go" button of this card(see chapter 6).

3.8 Other Cards

- Definitions and other formal stuff. It is accessible from the Go button under "Formal Lexicon Definition" (see chapter 17)

- Settings, Preferences and Paths for the lexicons, the word lists and the languages.

For futher details select the corresponding line of the "Go" button of this card (see chapter 12).

3.9 Other Useful technical details

"⌘ + space" shows (and hides) the menu bar of the HyperCard system. It is necessary for changes of the system only. Please be careful with such changes.

You cannot move the cursor with the arrow keys □ □ . In HyperCard they are used to move from card to card.

It is not necessary to save anything in fields when you add or delete something or when you edit rules. Hypercard saves permanently without explicit commands. Exception: You must store your completed entries explicitely from the entry card to the internal lexicon card and later on (after some days) from there to the external lexicon file.

To leave this program go back to the front card (click the home symbol) and click there the button in the left lower corner.

4 Practical Use of HyperLAT

4.1 How to Use HyperLAT

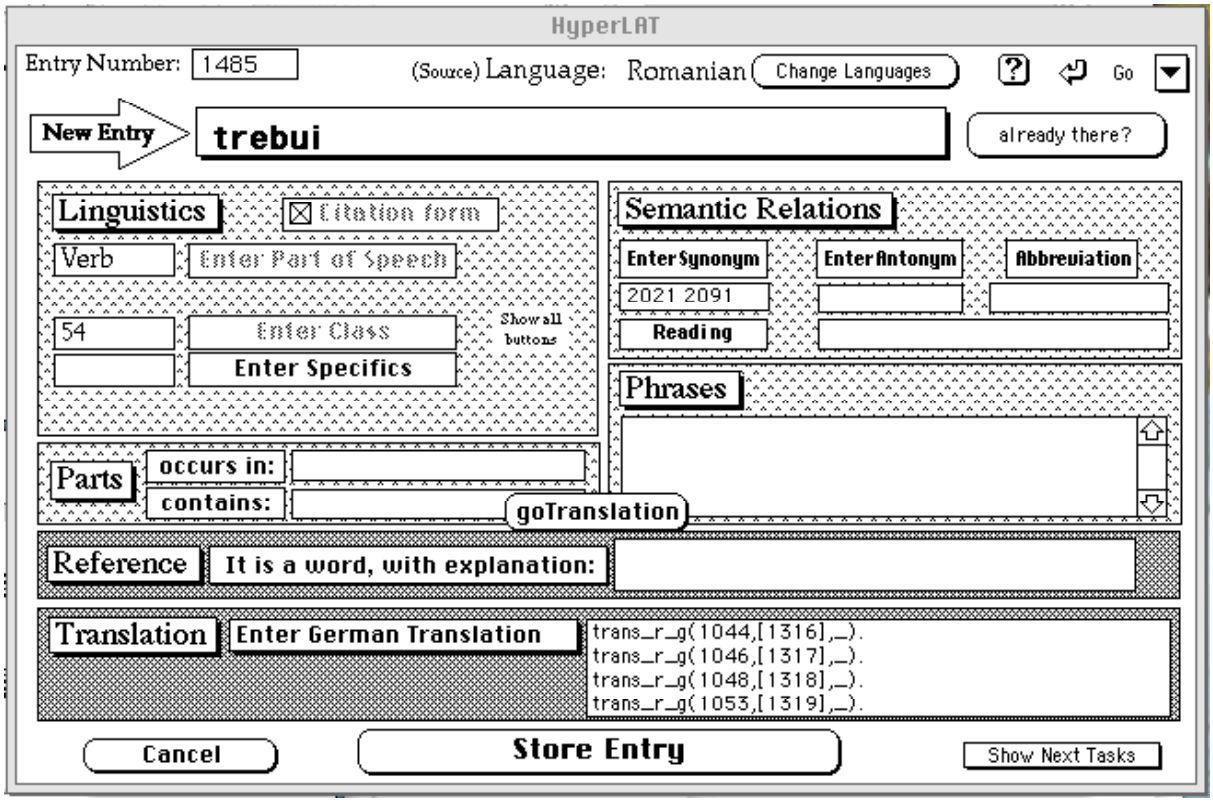

Figure 5: Acquisition card with Romanian setting

General remarks:

Access the card for entering new lexicon entries from the home page by selecting "Make New Entry" from the right button "Start Work". You can go to the home page by clicking "Welcome Card" in the popup menu of the "go" button.

Erasing and changing:
In case the user entered a wrong form in the main upper field of the Entry Card he/she should click the button "New Entry"or "Cancel" to reset all variables and specifications, which may have specifiend meanwhile.

If you chose wrong specifications from tables (part of speech or classes), you can easily overwrite the entry by clicking the corresponding button again.
Choices from the word list (for synonyms etc.) must be deleted manually from the field.

Read now the general infos about the acquisition card. You can go there by using the left arrow key.

4.2 Steps to enter new lexicon entries

Step 1. Language pair.

Set the languages to the desired language pair, if necessary. A field on top of the card indicates the source language. The target language (important only if you want to enter translations) can be seen from the title of the "Enter Translation" button in the lower right corner. Setting the languages will be done on the card "Paths and Preferences" if you use the button "**Set languages**".

Step 2. Word/term entry

Enter the (n-)word or (n-)term in the large upper field.

Step 3. Part of speech

Enter the part of speech of the word by clicking the rectangular button. A filed of choice will appear. Clicking on the correct line enters the standard abbreviation in the field next to the left.

Step 4. Gender (visible for pronouns only, otherwise hidden)

For pronouns you must click the "Enter Gender" button and select the correct gender from the small list appearing next to the button.

Step 5. Grammatical class (visible for simplex entries only)

for any word (except complex) you must choose now among alternative grammatical classes within the word classes (which you defined by the part of speech field). Appropriate fields for choice will appear.

Step 6. Syntactic structure (visible for complex entries only)

For complex entries you choose now the appropriate structure description from the field.

Step 7. Specifics

For many Parts of speech there exist exceptions which you may choose from the table. More than one symbol is allowed.

Step 8. Synonyms

Synonyms or antonyms (step 8) can be choosen from the Word List. You can choose as much words as you find synonyms. Their numbers are then put (by clicking in the line of the Word List) into the synonyms field ; the numbers there must always be separated by blanks.

If the word/term is not yet entered in the list, you can produce a short entry for that word/term by clicking "Entry Missing?". The following guided process allows you to enter the missing synonym /antonym in a preliminary entry . When the word occurs later as a "real entry" you can edit the short entry) (this updating is not yet implemented). The number of the new short entry is entered in the synonyms field.

Step 9. Antonyms

(see step 8: synonyms)

Step 10. Abbreviations

can be entered directly in the field by typing

Step 11. Reading

If a word has homonyms you can distinguis them by giving a meaningful word to remember which is which. It will be included in the word list and appear in the phrases segment of the lexicon.

Step 12. Compounds, complex entries and their parts

by clicking the buttons "contains:" and "occurs in:" you can choose - just as with the synonyms - contained elements and containing alements of the entry. Again, you can produce short entries for quick progress.

Step 13. Reference

If "It is a term" is checked, you can identify the graph or set of graphs which are a definition of the entry (not yet implemented)

If the entry is a word, you can enter a natural language explanation for that word, adding futher remarks etc.

Step 14. Translations

Here you may enter direct correspondencies to target language words. Click "Enter <langugae> Translation". You will see the translation card and can choose words from the wordlists there. Additionally you can add annotations on the translation card. The result will be saved directly from this card in its final form (it is shown afterwards on the card with your new entry).

If the target word is not there, you can produce an incomplete entry by clicking "entry missing?"

Step 15. Storage

clicking the middle button on the base line ("Store Entry") will

- enter all your specifications,
- concatenate them according to the lexicon format,
- show you the final entry,
- ask to you to edit this formula, if necessary,
- calculate the new entrie´s Id number,
- write the complete entry into the preliminary lexicon of the
- lexicon card,
- write all other lexicon parts (explanations, phrases etc. into
- the lexicon card,
- produce an abbreviated entry in the Word List, and
- clear all field for the next entry.

A temporal field will remind you that you saved this lexeme (in case you have been disturbed for a while). When you touch this field with the mouse, it will disappear.

(Step 16. Next entry

click the button "New Entry" in the base line to enter the next word, and continue by step 2. Otherwise some fields may not be cleared by the previous values and could lead to non-grammatical entries.

(Step 17. Check lexicon entry)

If necessary, you can check the new entry again by clicking the button "Show Lexicon" down on the right. This will show the lexicon card. Editing there will be stored automatically, as anywhere in Hypercard. Be careful when editing entries, they really might be correct!

Attention!
the new entry is now stored only in the **internal lexicon of HyperLAT**.
No "DBR-MAT Lexicon" entry has been produced so far!

Only if you click on the button **"Store Lexicon to File" on the lexicon card**, this lexicon is written into a file "D-BMAT Lexicon#" (where # = an automatically increased number). These files are created for technical reasons in the same folder with Hypercard).

Read the Intro to the Lexicon Card for further information (see chapter 8)

(Step 18. Update the lecicon file)

If you click on the button "Store Lexicon to File" on the lexicon card, this lexicon is written into the file "D-BMAT LexiconT" (which must be in the Hypercard folder).

5 The Acquisition Card

(for surface picture see
Figure 2 and Figure 5

5.1 General remarks

On the "HyperLAT Acquisition Card" the user is supposed to enter new entries with all their obligatory and additional features. If you click here, you will see the acquisition card for a short time and come back to this card automatically.

Please keep in mind, that you have to set the language pair when other languages (and other character sets) are under work. A field on top of the card indicates the source language. The target language (important only if you want to enter translations) can be seen from the title of the "Enter Translation" button in the lower right corner. If you click the button **Change Languages** you will be sent to the "Paths and Preferences" card to define languages consistently.

In this introduction we distinguish the following entries to the acquisition card:

a) Single entries:

- word = everyday word, which can be simplex ("fist") or compound ("furthermore"); the semantics of words is empty or can be an explanation text only.
- term = technical language expressions ("carburator"), simplex or compound ("doubleclick"); the semantics of terms is their link to the knowledge base of conceptual graphs.

b) Complex entries

- n-word = a word with more than one element ("monster pizza service"), separated by blank(s)
- n-term = terms of more than one element ("very large scale intergration", VLSI).

Entries of all these categories and their inflected forms occur in the lexicon. They have different guided entry procedures each.

5.2 Functionality

The user types in either a word or a term, both, simplex or compound, simple or complex.

The button "**Already there**?" allows for checking whether a word is already in the lexicon. If the word is really new, the button has "no!" result. Otherwise the entry is shown on the word list card.

Successively he/she can define the word ***linguistically*** by entering

- the part of speech,
- the gender, if applicable,
- a grammatical/inflectional class (for single words/terms or
- a syntactic class (for complex entries, i.e. entries with more than one separated element, respectively).

Only applicable buttons are enabled by the system. For each entry step a specification field (offering the possible choices) is provided, to ensure consistent and homogeneous entries. The necessary fields and buttons are enabled, disabled, shown or hidden according to the grammar.

Under ***semantics*** synonyms, antonyms and abbreviations can be entered. The requested reference to other words is provided by selection from a "Word List", a summary of all entries and their identification numbers. Additionally you can enter one word to sketch the reading (meaning) of a word for disambigua- tion. The reason, why: In the word list you may find the line "bank 789" but you do not know whether it is the financial bank or the park bank. The entry "reading" is only used on the word list

For compounds or complex entries you can mark which parts the lexeme **contains** and in which lexemes the entry **occurs**. Again the choice is made from the "Word List".

If you are sent to the "Word List" card and you discover that a word is not yet in the Word List (hence, the lexicon), the opportunity is given to enter this word immediately, because synonyms and antonyms, if not necessary in the lexicon otherwise, can be stored as short entries without detailed linguistic information. After this interruption the user can continue entering the original lexeme.

Reference provides two alternatives (shown mutually excluding): If the button shows "It is a term…", you can enter (in future versions) the corresponding graph id (not yet implemented). If the button is"It is a word…" (it is an everyday word), you can enter a free text explanation in a text field.

The header field in ***Enter ...Translation*** changes its title according to the specified target language (a simple way to see, what is defined). In the text field you will see the formal entry after the definition on the translation card (see its help card under "Go").

If you click "Enter ... Translation" you will go to the translation card

The other buttons on the card allow

- to clear all entries for a ***new entry,***
- to ***cancel*** the entry process at any time,
- to ***store an entry.*** First you will see the completed formal entry in its final form,
- to show or hide additional ***fields*** (like a note pad "To do")
- to check whether an entry is already in the lexicon "(already there?")

6 Translation

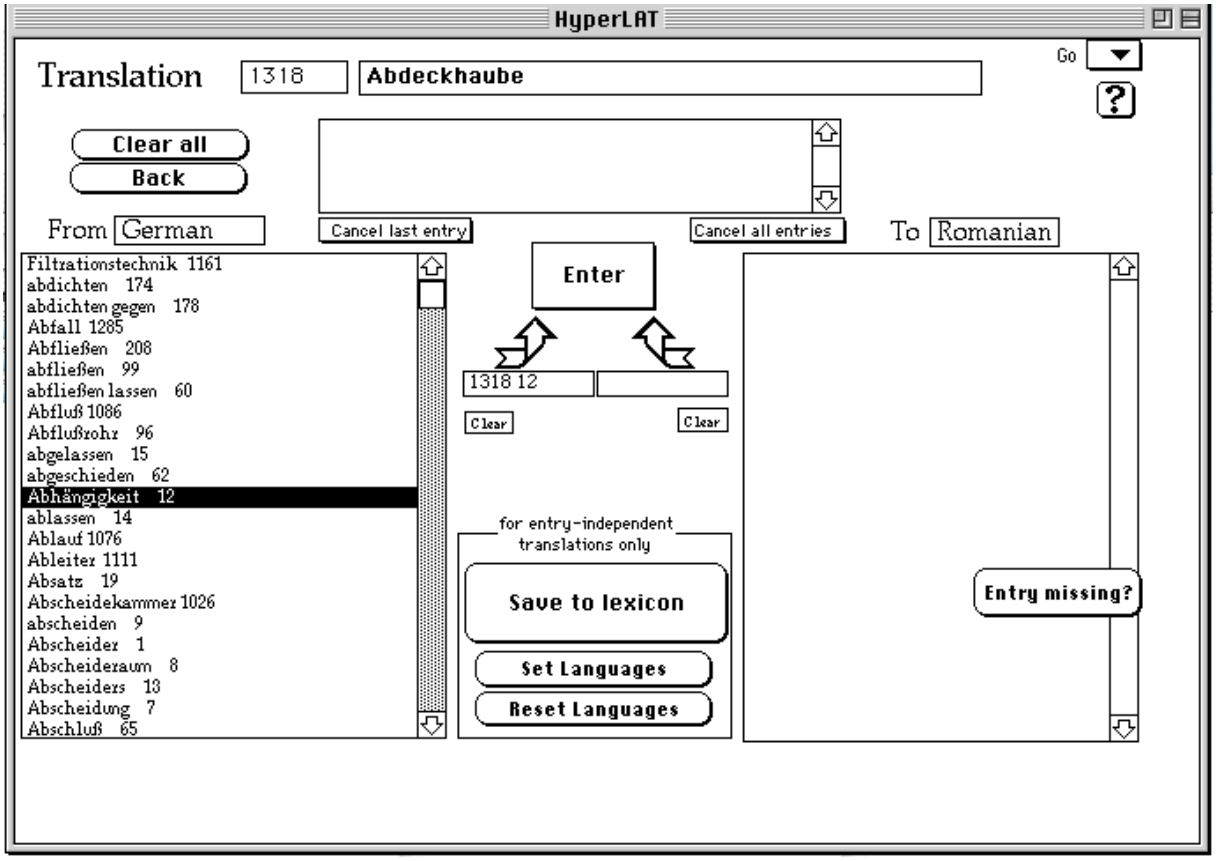

Figure 6: Translation card

The translation card allows for correlating two or more Ids of words from different languages. For this purpose the word lists (abridged versions of the full lexicons) of the chosen language pair are shown.

If you click here, you will see the translation card for a short time and come back to this card automatically.

The language pairs are predefined (on the card "Paths and Preferences") for the work at hand on the Acquisition Card. Therefore in most cases you will come to the translation card from the Acquisition Card (where you clicked "Enter Translation"). You will see the word you are about to enter in the upmost field together with its Id number. This number is already entered in the source language field and will be used in the translation entry as source language word.

To write a translation entry follow these steps:

1. select a target word from the right list. This will put the Id of this word into the small field left to the target word list. If you want to enter a synonym, which corresponds to the same source, click again in the target word list. The Ids will appear in a line.
If the intended word is not yet in the word list, you can produce an incomplete entry for this word by clicking "entry missing?"

2. click **Enter** to make a real entry out of the Id list and to go back to the Acquisition Card, where you will find the short formula in its field.

You may, however, decide to enter translations with other pairs than those predefined, and without coming from the Acquisition Card. This is possible and supported by the system.

Do the following;

1. check (or change by "**set languages**") the language pair to get the correct word lists to choose from. The defaults for the Translation Card are the languages defined on the card "Paths and Preferences". These defaults can be reset (by "**reset languages**") after work with another language pair.

2. select (click) a source word from the left list. This will put the Id of this word into the small field right to the source word list. If you want to enter a synonym, which shall have the same target, click again in the source word list. The Ids will appear in a line.

3. select a target word from the right list. This will put the Id of this word into the small field left to the target word list. If you want to enter a synonym, which corresponds to the same source, click again in the target word list. The Ids will appear in a line.
If the intended word is not yet in the word list, you can produce an incomplete entry for this word by clicking "entry missing?"

4. click **Enter** to make a real entry out of the Id lists. This entry will appear in the topmost field. You cannot edit these entries; in case of errors use the small **Cancel** buttons.

5. Proceed with the next translations until a reasonable set of translations is written.

6. Save the set of new entries to the main lexicon by clicking "Save to lexicon". The field will be empty afterwards. You can check the result of your work on the lexicon card. NB: Entries will be saved in the lexicon under the source lexicon (the first variable of the prefix "trans_x_y", i.e. the _x lexicon), therefore

7. If you change the language pair underway you will be asked to save the existing entries first.

The translation entry is saved to the lexicon directly from this card, even if you came from the Acquisition Card. However, the translation entry is shown there on the card with your lexicon entry under work.

7 Word Lists

The word list is an abridged form of the lexicon entries consisting of

<Lexicon item> <Id number> (<Reading>)

7.1 Where is the word list used?

If you refer to other words in a DB-MAT lexicon entry, HyperLAT will enter only the Id (identifier) of the corresponding word. Links to other words are needed when you specify synonyms, antonyms, citation forms for inflected words, contained words, containing words, or translations. To avoid tedious search for Ids in the full lexicon, an abridged form of the lexicon (containing only the words and their Ids) is kept by the system. From there you simply click a word and HyperLAT will include its Id in the entry formula.

If you click here, you will see the German word list for a short time and come back to this card automatically.

If a word that you need as reference is not yet in the lexikon (hence, not in the word list), you may enter this word immediately as a short entry (by clicking the button "Entry missing?" on a Word List

Card) without detailed linguistic information. After this interruption you can continue entering the first lexeme.

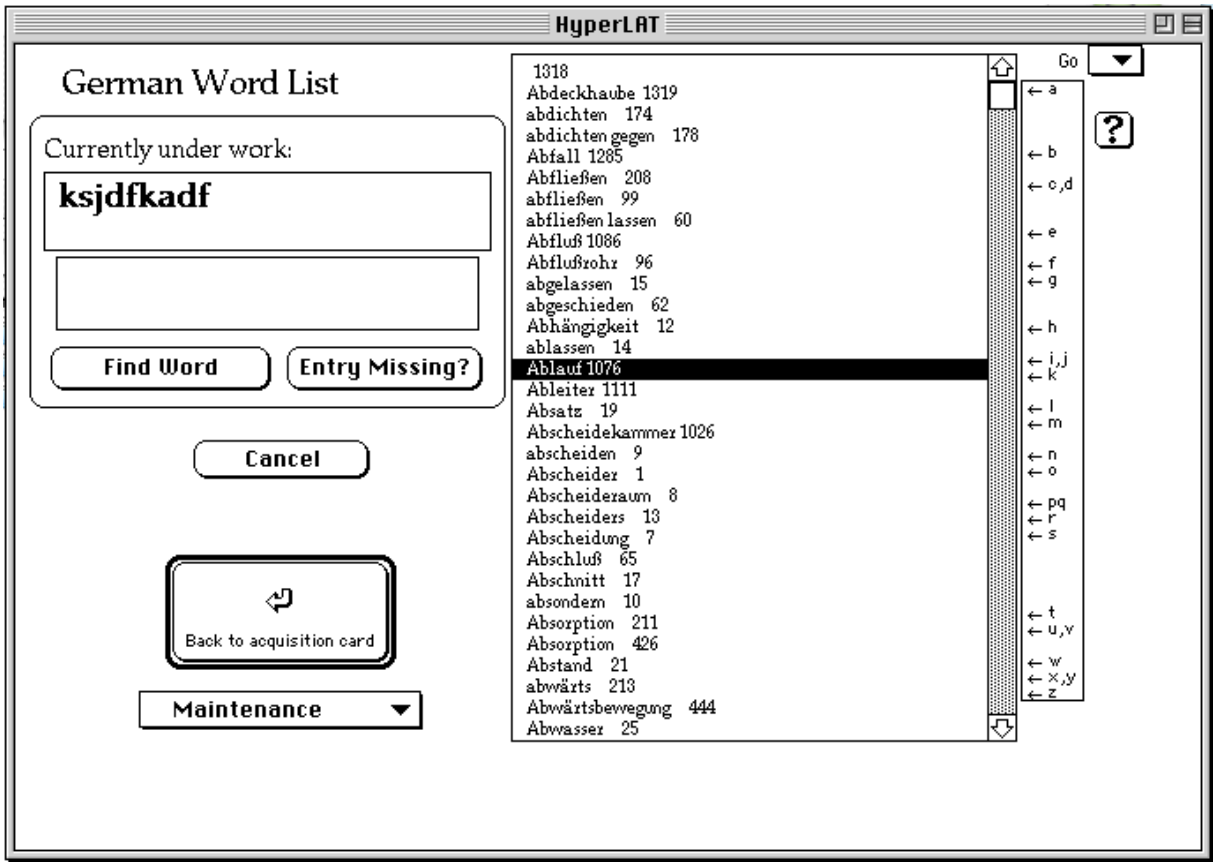

Figure 7: The word list card

Example:

You are about to enter the word "near". After defining other linguistic features you want to enter an antonym and you decide to refer to the word "distant". Your search in the word list, however, turns out to fail: the word "distant" has not yet been entered. You click the "Entry missing?" button and you are brought to the acquisition card automatically. There enter "distant" in a new, smaller field appearing on the acquisition card. After saving this intermediate and incomplete entry the new id number of "distant" is already entered into the field "Antonyms" and you can proceed with further definitions of "near".

7.2 Where does the word list come from?

Every new word acquired by HyperLAT is automatically appended (in an abbreviated form) to the internal word list of that language.

You can always produce a new word list out of an existing full lexicon by following steps (1) to (3):

cutting the language specific prefix "lex_entry_x" off the DB-MAT Lexicon entries,
putting it into a new file "Lexicon_x" (x being the language abbreviation g, b, r, p, f or e) with MS Word 6.0, and
starting the Word macro "MakeWordList" (if you do not have it, send a mail to "vhahn@informatik.uni-hamburg.de").

The result is a new word list which should be saved in an external ASCII file "Wordlist_x", where x is a language id . It can be loaded into HyperLAT by a path definition on the card "Preferences and Paths".

The word list must always have two or more blanks between the word and the id. No, I don't know why. The last line of a word list must be empty. No, I don't know either.

The word list card allows for simple maintenance: You can print the word list, sort it aphabetically or start loading a new word list (via "Preferences and Paths").

Attention:

1. Do not produce or load a word list from a lexicon file, which does not contain the last entries made by HyperLAT! Remember: HyperLAT does not save the new entries automatically to the external lexicon file!

2. If you delete an entry from the lexicon, the corresponding word must be immediately deleted in the word list too. Otherwise future entries can use a non-existing id.

8 Internal Lexicon Card

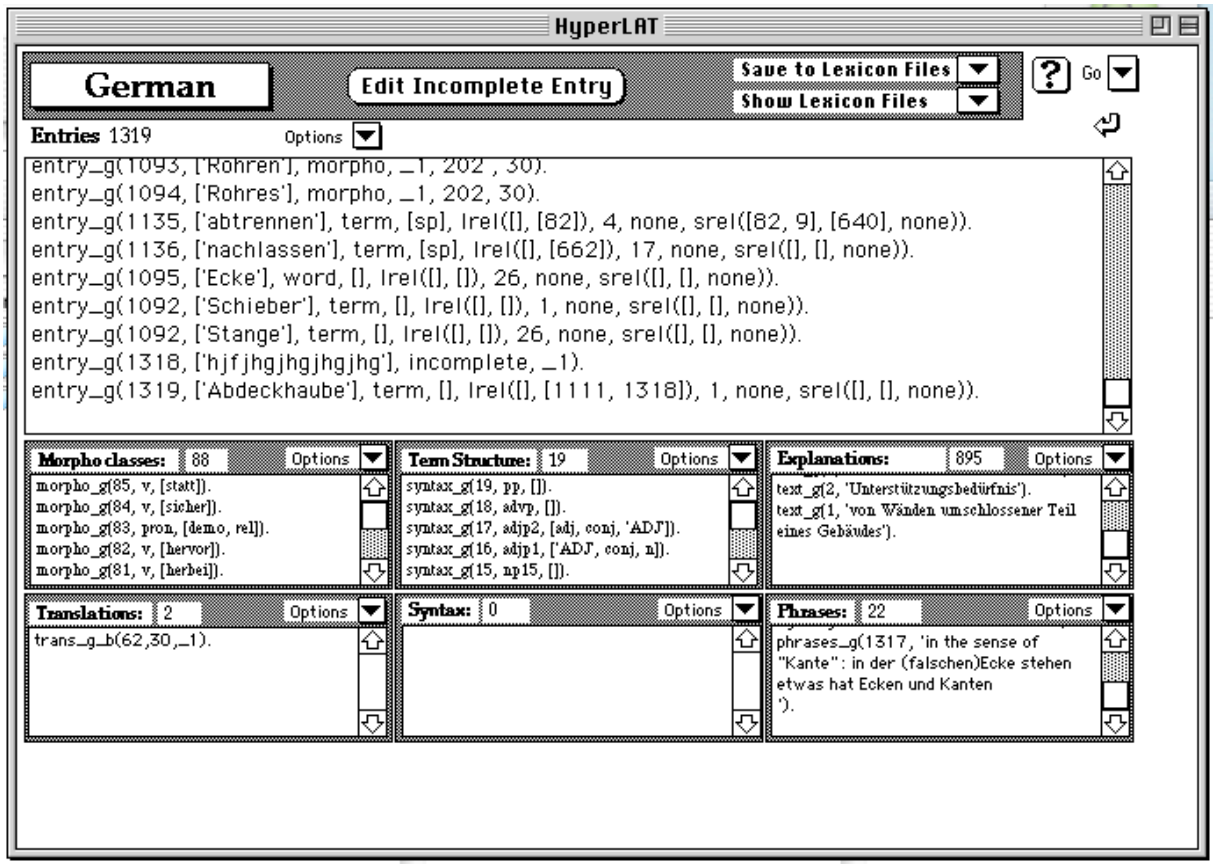

Figure 8: The lexicon card

8.1 The lexicon card contains in 7 fields:

The lexical entries (DB-MAT prefix: "entry_x") of words entered by HyperLAT,
the explanations (DB-MAT prefix: "text_x") attached to words (in contrast to terms),
3. the translations (DB-MAT prefix: "trans_x") defined among two entries,
the morphological classes (DB-MAT prefix: "morpho_x") consisting of Ids and the rule code) applied in this language, and
the syntactic NP rules (DB-MAT prefix: "struct_x") consisting of Ids and syntactic descriptions for complex terms in this language
the syntax information, mainly used by Bucharest (DB-MAT prefix: "synt_x") defined for words,
the phrasal information (DB-MAT prefix: "phrase_x") defined for words to indicate their collocations

On the lexicon card you can check new and existing entries of the lexicon. Note: Changes in any file of this card will be stored within Hypercard automatically! Changes, however, will not change the **external** lexicon file.

Please note: New entries made by HyperLAT are first stored only on the **internal lexicon card of HyperLAT**. No new entry in the final "D-B MAT Lexicon" will have been produced so far!

8.2 The HyperLAT lexicon card and the DBR-MAT lexicon

Only if you click the button "Store to Lexicon Files" on the lexicon card, the HyperLAT internal lexicon parts you specify are written definitely into the files specified as the lexicon files on the "Preferences and Paths" card. When storing the lexicon you will be asked by HyperLAT, whether the stored entries shall be kept in HyperLAT. If you keep them make sure that the old entries will not be stored a second time with the next new entries (you will be warned).

The internal lexicon can remain within HyperLAT for a long time without being stored to the DB-MAT lexicon (it may contain up to 30 000 chars).

You can inspect the "real" external lexicon files itself by clicking the button "Show Real Lexicon File" .

8.3 Incomplete entries

The lexicon contains partly incomplete entries without any linguistic specification. They were produced, when you want to refer to another lexical entry, which is not yet in the lexicon. This happens, when you were just about entering a new word, and can't find the synonym / antonym in the word list to click it into the corresponding field in the Acquisition card. Technically speaking, you decide to press the button "Missing entry?" on the "Word List" card (click here to go to the "Word List" card) and produce a preliminary entry to reserve an Id for it. Incomplete entries consist of the normal prefix (entry_x), the id, the word itself and the marker "incomplete".

By clicking at Edit incomplete entry you can expand incomplete entries to full entries. Select an entry and click the button. You will be sent to the acquisition card with the appropriate parameters already set.

This procedure ensures that

(1) the Id number of the incomplete entry is kept (although you produce a "new" entry) and
(2) no changes in the top numbers are made.

8.4 Other fields and buttons

The top entry numbers shown on the lexicon card over each field (the highest number of this sort of entry) mirror the corresponding fields on the card "Paths and Preferences" and cannot be edited on the lexicon card.

8.5 The Option button

Show in Wide Window
displays the data on the whole card and may be necessary to inspect a greater number of entries. A button appears on top of the card to shrink the field to its old size

Show Save History
displays a field with the dates of previous save activities. A button appears to hide this field again.

Delete Last Line
deletes the last line of this filed and subtracs 1 from the number (on all occurences of on the grammar card and the preferences), This is necessary for the bookeeping of the DBR-MAT system. If you delete the last line of the lexicon entries, the corresponding lines under "text", "phrases" and "translations" will be deleted too.

Sort Field
sorts the field (in a sometimes not predictable way, sorry)

Lock Field
protects all fields and should be clicked after any manipulations on fields

Unlock Field
allows for writing or removing lines in all fields

remove +++ lines
you may want to remove the markers for saving activities

Print this Field and Print all Fields
are obvious actions. However, remember, that you are only printing the intermediate HyperLAT results, not the DBR-MAT lexicon. This is only in the file (unless you never saved anything)

9 Grammar Maintenance

9.1 General information

(a chapter "Troubleshooting" is at the end of this field)

On this card you can edit those tables which are shown to the user temporarily to choose from on the Acquisition Card. Whenever you have to fill in something on the Acquisition Card, the system will show you the possible choice in text fields. It is

a table for the parts of speech,
the morphological tables for each part of speech,
specific features of single words, and
the constituent patterns of terminological phrases (xomplex terms).

On the grammar maintenance card you can edit the contents of these fields. The grammatical rules will become numbered entries in the lexicon.

If you click here, you will see the German grammar card for a short time and come back to this card automatically.

Additionally this card shows the highest (i.e. latest)
- entry number,
- morphological rule number, and
- syntax rule number.

All editing must keep consistency between the rule Ids, the top numbers and the rules themselves.

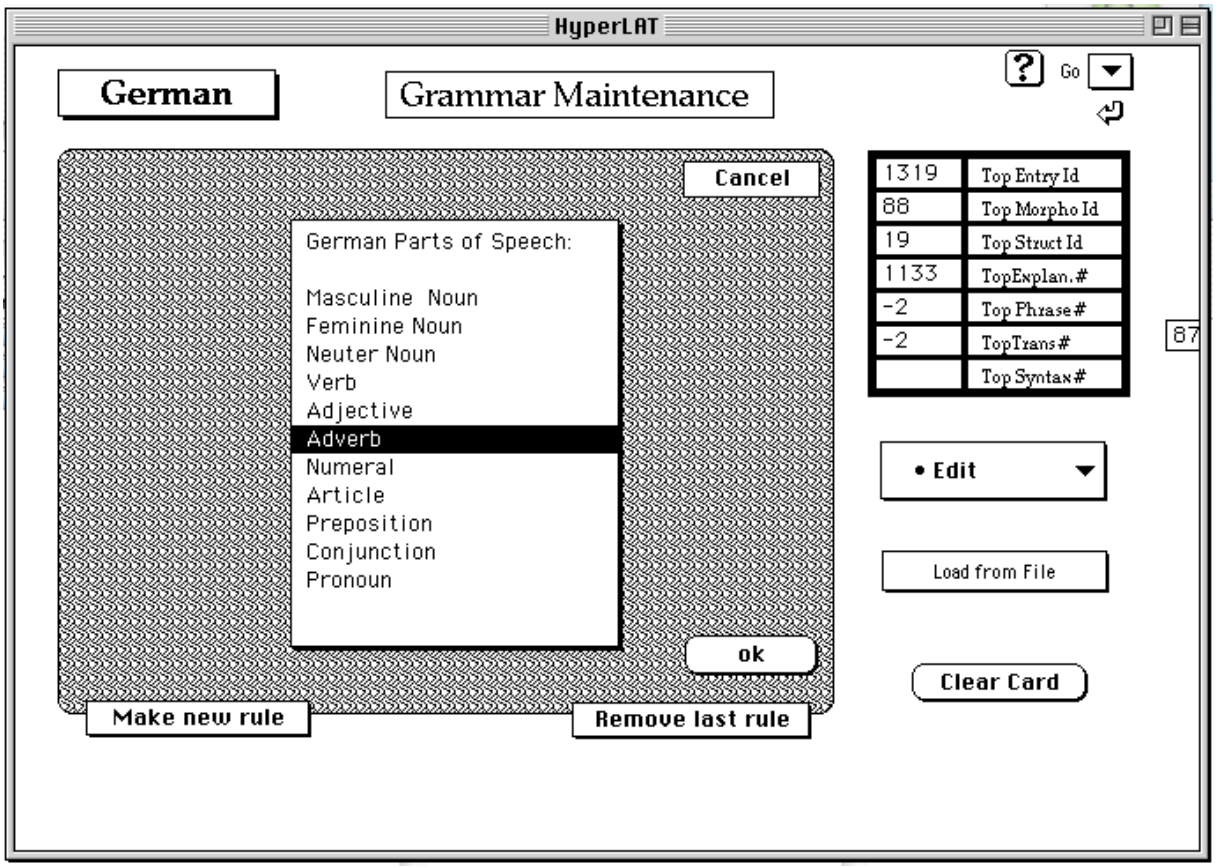

Figure 9: Grammar maintenance card

The Grammar Maintenance Card allows you to edit all entries, to add new rules or classes or to remove them. In the Acquisition Card this is not possible and the text of the fields is locked there (to enable clicking) and to keep the consistency with the Ids.

The existing fields in HyperLAT contain all the rules of the existing lexicon and some additionall proposals for further rules, extracted from a German grammar. Esp. for the words other than nouns, HyperLAT proposes a set of maybe useful classes. These rules and classes do not have any Id. When clicked they produce an empty entry. If you want to use them, produce regular entries.

9.2 How to edit grammar rules:

Click the buttons for the corresponding sort of editing (morphology, term structure, word specifics, or the parts of speech). In case you changed your mind while editing you can always close the fields by clicking "Cancel" or "Clear Card". If you start with a new language, define the parts of speech first.

Change entries

You can change all existing lines with the editor as long as you keep the number of rules or classes. i.e. you must not type in or erase Ids!

Add rules

Use the button "Make new rule". It produces a new Id, increases the corresponding top number, and puts the Id at the beginning of a new line in the open field (if not, edit it). This is necessary to recalculate the Top Morpho/Syntax Ids and to keep track of the entries in the lexicon.

9.2.1 Cancel rules

If you want to remove the last rule, use the button „**Remove last rule“**!

Because: The last rule has the "Top Syntax Id" or "Top Morpho Id". If it is deleted, the corresponding Top Id must decreased by 1 (which is done if you use the button).

Rules or classes with other than the top number can be erased from the table without any problems. Gaps in the numbering are allowed. If it is inevitable to erase rules, remove them
- from the field at hand,
- from the lexicon card (for morpho and struct rules)
- from the lexicon itself (for morpho and struct rules)
- from the data files (for the specifics and the parts of speech)

The new entries are saved (as always in HyperLAT) locally as you type them. No explicit "Save" is necessary.

9.2.2 Parts of speech

Part-of-speech entries are a bit tricky, because if you add a new part of speech, this means

• a new line must be added in the field "Part of Speech" which appears on the Acquisition Card when you press "Enter Part of Speech". OK, but:
• the next step on the Acquisition Card is choosing from the classes of this new PoS.
Therefore the Grammar Card creates a new empty specification field for each new PoS you enter.

9.3 Trubleshooting

9.3.1 Remaining fields

If a field remains on the screen and does not hide when you click "Clear Card" do the folowing:

Either

Choose from the Edit button the corresponding "Grammar Class" or "Word Specifics" to which the remaining field belongs.

Click "Clear Card".

Or

1. Choose "Tools" in the menu and select the field tool on the upper right corner.
   doubleclick the remaining field. You will see the Id number of this field.
2. Open the message box by pressing the apple key and the "m" key together.
3. Type "hide cd field id <nn>", where <nn> is the Id number of the remaining field.
4. Click the close square in the upper left corner of the message box.

9.3.2 Wrong fonts in fields?

1. Type + spacebar to show the menu bar
2. Choose the field tool from the menu (rightmost in the upper line)
3. Doubleclick the field. the Info window appears
4. Click the "Text Style…" button and define the font and size

5. Close the Info window
6. Choose the browse tool from the menu (the "hand", leftmost on the upper line)

9.3.3 Fields too large or in the wrong place?

1. Type + spacebar to show the menu bar
2. Choose the field tool from the menu (rightmost in the upper line)
3. Move the field with the mouse (button down) or
4. Take the field on one corner and pull it smaller with the mouse (button down)
5. Choose the browse tool from the menu (the "hand", leftmost on the upper line)

10 Saving Grammar Data

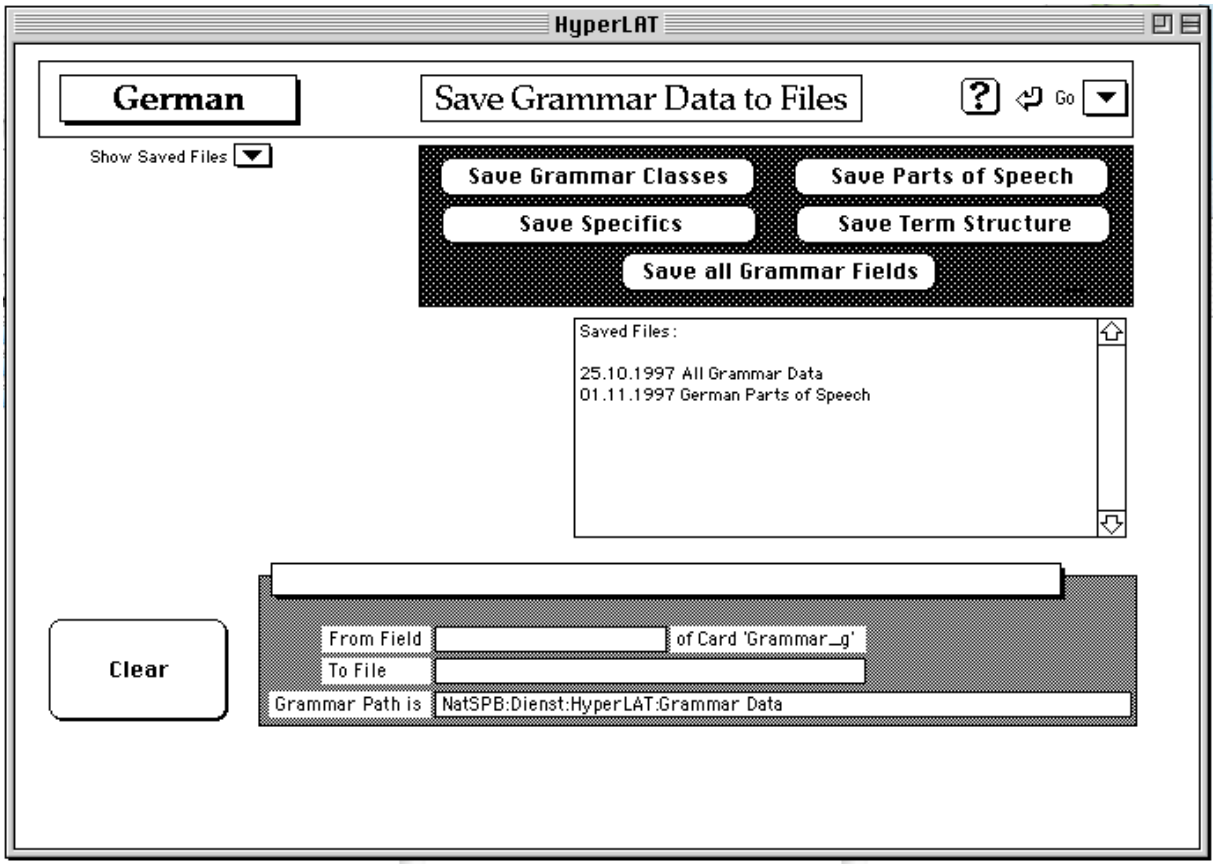

Figure 10: Saving grammar data card

From this card you can save grammatical data to external files (those in the HyperLAT folder). Every field (for parts of speech, for maculine nouns, for adverb specifics etc.) is saved to a separate file to allow for reloading.

You can save:

1. Parts of speech
2. Grammar classes

3. Word specifics
4. Term structure or
all of them

If you choose 2. or 3. you can choose from a list those fields which should be saved. On the lower part of the card you see which action will be performed when you press "Save".

An additional field shows all save actions with its date.

11 Loading Grammatical Data

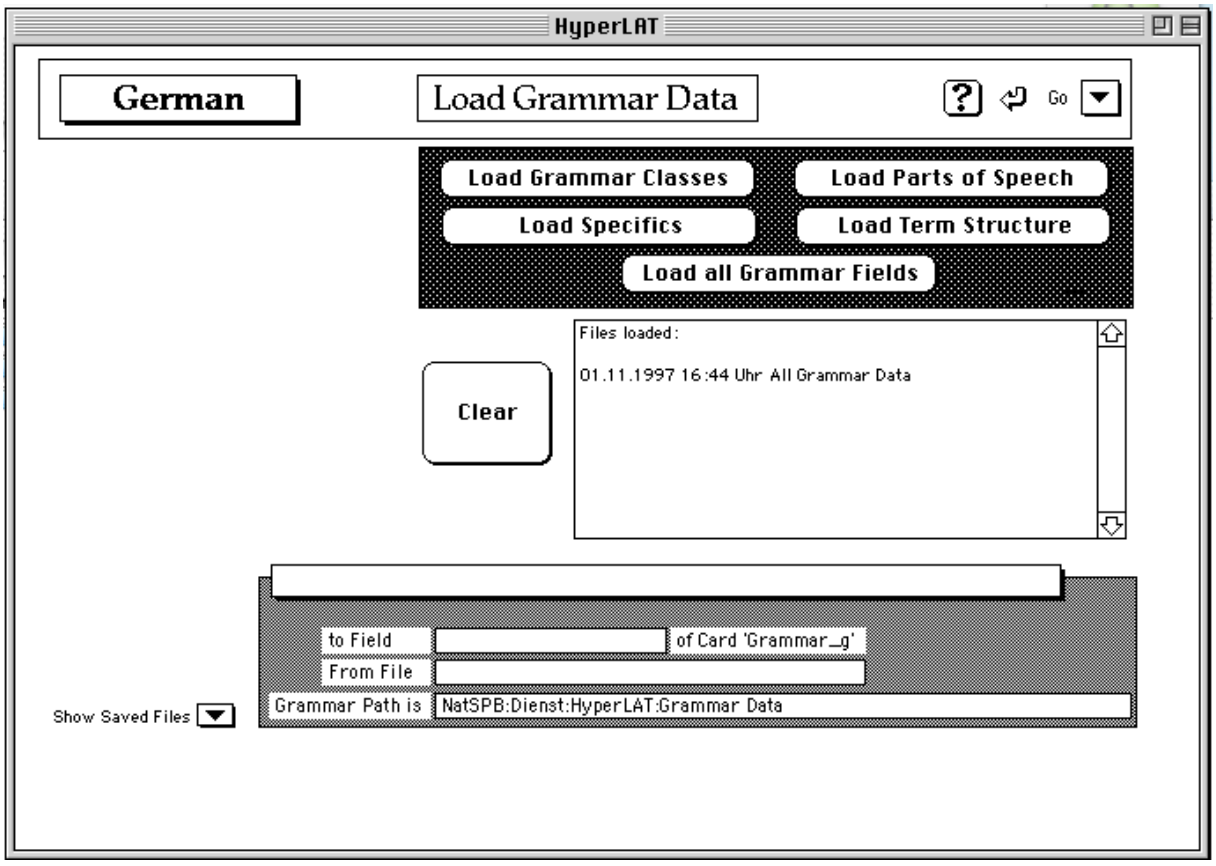

Figure 11: Loading grammar data card

From this card you can load grammatical data from external files (those in the HyperLAT folder) to the HyperLAT fields on the Grammar Maintenance Card. Every field (for parts of speech, for maculine nouns, for adverb specifics etc.) is loaded separately.

You can load:

1. Parts of speech
2. Grammar classes
3. Word specifics
4. Term structure or
all of them

If you choose 2. or 3. you can choose from a list those fields which should be loaded. On the lower part of the card you see which action will be performed when you press "Load".

An additional field shows all loading actions with its date.

12 Paths and Preferences

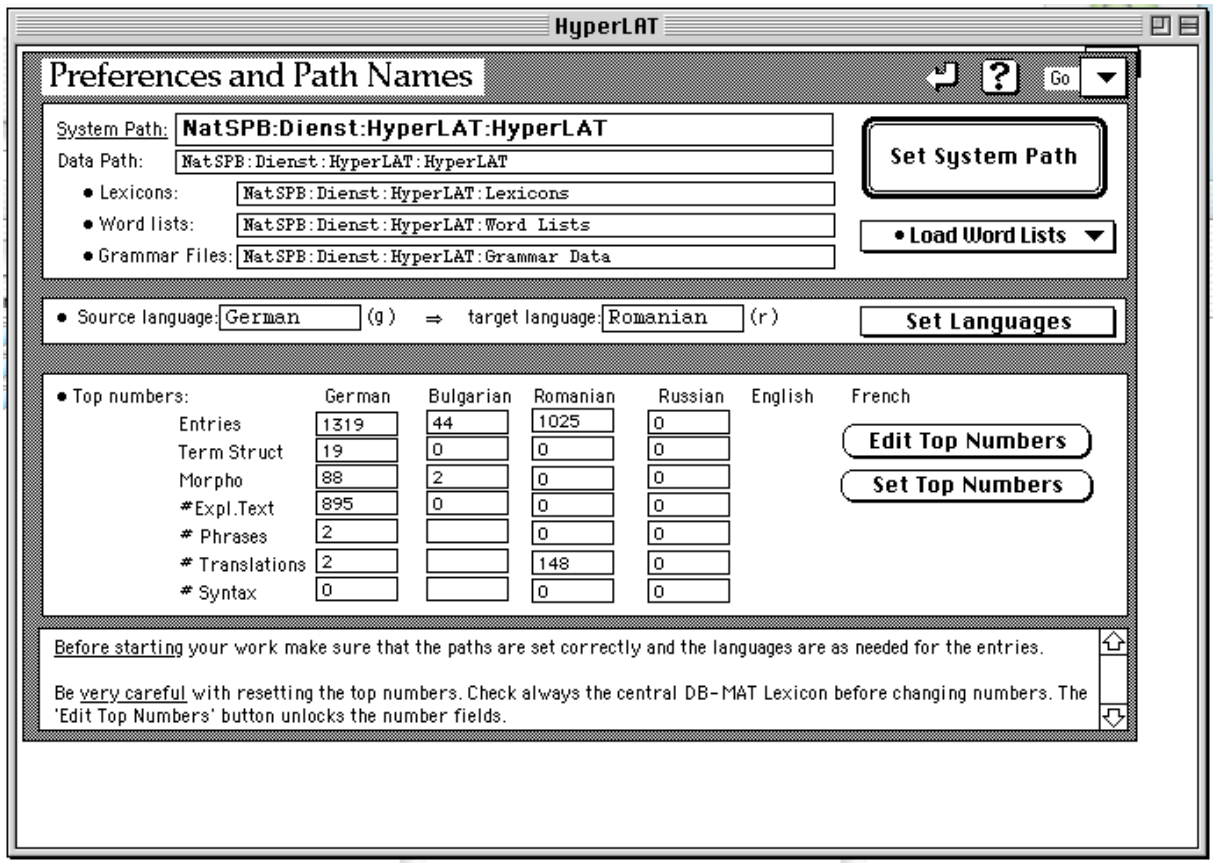

Figure 12: Paths and preferences card

On the "Paths and Preferences" card the user

- must define the system path,
- can change the language pairs, and
- can (re-) set the top numbers of entries

12.1 System Path

To set the system path is necessary to find the lexicons, the word lists and the grammars when saving entries. When you have installed the system the paths must be set in the beginning. This information is kept until you change the position (hence the path) of the system.

For more information about the HyperLAT folder click here.

12.2 Language Pairs

You can change the language pair for the current work. When entering translation equivalents, you can change temporarily the languages for single entries without changing the general language setting of the system.

12.3 Top numbers

The DBR-MAT lexicon uses the numbers of the highest entry of each type (lexicon entry, term structure, morpho rules, etc.) for several reasons. It is not the abolute number of entries, but the highest Id number in use. This information is stored in the lexicon (at the very end). The bookkeeping of the top numbers is done consistently within the HyperLAT system.

Additionally HyperLAT sums up the number (#) of explanations, syntax entries, phrases etc.

This makes changes of the lexicon a bit difficult, because each time

- you add an entry you have to add 1 to the top number, and each time
- you delete the last existing entry, you have to decrease the top number by 1.

Sometimes you made some test entries and want to delete them by hand and afterwards the top Ids are not any longer consistent. Note that gaps in the numbering do not disturb the system. Only the number of the joungest entry (the "top Id") is important.

The numbers (#) of the other entries, like explanations or phrases are only used informally to estimate the size of the lexicon. Inconsistencies do not produce crashes.

Therefore the entries can only be deleted on the lexicon , one by one, keeping the Id numbering consistent.

You can set the numbers in HyperLAT on the "Paths and Preferences" card. These numbers are propagated consistently throughout the system.

The fields with the top numbers are normally locked. The button "Edit Top Numbers" unlocks them for changes. By hitting "Set top numbers" the fields will be locked again and the new numbers are propagated throughout the system on all other cards.

13 Troubleshooting

1. The window is always at the left upper corner, which is a bit inconvenient!

Solution: open the message box (apple+m) and type
set the loc of window "Hyperlat 0.8.5" to 100,100
(change the name of the system as necessary)
then you can type to the msg box "show titleBar"
which allows you to move the window around
with the bar

2. The window is too big

Solution: open stack info (under Objects), click "Resize" and
set the card size to "Powerbook"

3. Inspecting real lexicon files may cause the message
"Can't open this application" (or similar).
In most cases the reason is unsufficiently allocated space for the program. Quit the application, if it is actice and click once the icon of the application (e.g., SimpleText). The icon gets black, then type apple+I (= I nformation). You will see the allocated space. Set the preferred space (last line) to a higher value (for SimpleText, e.g., 1200 K). Close the field (left upper box) and try again with HyperLAT

14 Technical Information

14.1 General Information

This programm is written as a hypertext with the system HyperCard and its script language HyperTalk. If you are not acquainted with HyperCard, please read the first chapters of the HyperCard manual first, or go to the HyperCard Help file of HyperLAT (click here if you want to see it).

HyperTalk allows for scripting the HyperCard objects. The distributed HyperLAT may, however, come as application, not as the source program. In the application you cannot edit the scripts behind the objects.

The path name of the HyperLAT system must contain only single names! (no "HyperLAT System" or something like that).

To make sure that the internal relative path names remain the same, the following files are necessary in the HyperLAT folder (and must be in the correct subdirectories, as described here):

15 The HyperLAT Folder

The path name of the HyperLAT system must contain only single names! (no "HyperLAT System", "HyperLAT 0.8.7" or something like that).

In the HyperLAT folder the following files are necessary (and must be in the correct subdirectories, as described here):

1. *HyperLAT*. HyperLAT itself

2. *Lexicons*. The DBR-MAT lexicons for the languages each in the folder "Lexicons"
   - 2.1 *Lexicon_g*
   - 2.2 *Lexicon_b*
   - 2.3 *Lexicon_r*
   - 2.4 *Lexicon_p*

3. *Word Lists*. The word list files for the languages each
   - 3.1 *Wordlist_g*
   - 3.2 *Wordlist_b*
   - 3.3 *Wordlist_r*
   - 3.4 *Wordlist_p*

4. *Grammar Data_x*. The grammar data in separate files for each language

   - 4.1 *Grammar Data_g* which contains
     - 4.1.1 *Grammar Data_g*
     - 4.1.1 *Parts of Speech_g*
     - 4.1.1 *Term Structure_g*
     - 4.1.1 *Word Specifics_g*

   - 4.2 *Grammar Data_b* which contains
     - 421.1 *Grammar Data_b*
     - 4.2.1 *Parts of Speech_b*
     - 4.2.1 *Term Structure_b*
     - 4.2.1 *Word Specifics_b*

4.3 *Grammar Data_r* which contains
- 4.3.1 *Grammar Data_r*
- 4.3.1 *Parts of Speech_r*
- 4.3.1 *Term Structure_r*
- 4.3.1 *Word Specifics_r*

4.4 *Grammar Data_p* which contains
- 4.4.1 *Grammar Data_p*
- 4.4.1 *Parts of Speech_p*
- 4.4.1 *Term Structure_p*
- 4.4.1 *Word Specifics_p*

5. *New in Version x* A file with comments to the latest version

6. this file.

Please copy always the whole directory!

Remember that all files save by appending the new material after the old one. So the files increase and you should clear them from outdated material every half a year...

16 Info on the DBR-MAT System

The DB-MAT system is a Machine Aided Translation Tool for technical translations from and to German, Romanian and Bulgarian. One of the most innovative features of the system is its knowledge background. A formal representation (Conceptual Graphs) allows for domain clarification queries. Independently of the lexical material the user can select chunks of text from one of the text windows and ask from a menu questions like: "What is X?" or "Who does Y?". The ansers use the CG structure (the statements and a conceptual hierarchy) and generate natural language answers from the collection of graphs which have been found.

The lexicon is used (among others) to

- access the knowledge base from the selected text,
- verbalize the answers (extracted from the graphs),
- look up translations,
- reduce inflected forms to citation forms,
- connect single word to multiword items,
- refer to synonyms and antonyms,
- refer to abbreviations,
- activate figures and pictures

...

(go back the the main info card by clicking the "General Info" button).

17 Lexicon Entries in a BNF-similar style (Version 3)

terminal elements are italicized

```
<LEXICON>                ::= one or more <LexiconLine>

<LexiconLine>            ::= {<Entry>, <Trans>, <Text>, <MorphoRules>,
                             <StructRules>, <SyntaxRules>, <Phrases>,
                             <SemRules>}
<Entry>                  ::= entry_<x>(<EntryId>, <EntryBody>, <Type>).
<x>                      ::= a character, the abbreviation for a language
                             (values:b = bulg, g = german, r =romanian, p
                             = russian, e = english)
<y>                      ::= a character, the abbreviation for a language
                             (values: b = bulg, g = german, r =romanian,
                             p = russian, e = english)

<EntryId>                ::= number of entry, unique for each language,
                             attached by the HyperLAT system
<EntryBody>              ::= ['< a lexeme>'] , if more than one: in the
                             brackets (remark: any component of a complex
                             term is quoted)

<Type>                   ::=  <Term>| <Word>| <MorphoType> | <Incomplete>
<Term>                   ::=  term, <Annotation>, <LRel>, (<MorphoId>,
                             [Specification]), <StructId>, <SRel>.
<Word>                   ::= word, <Annotation>, <LRel>, (<MorphoId>,
                             [Specification]), <StructId>, <SRel>.
<LRel>                   ::= lrel (<ContainsList>, <IsContainedList>)
<ContainsList>           ::= [number] | lists of Ids in [], if more: sepa-
                             rated by comma
<IsContainedList>)       ::= [number] | lists of Ids in [], if more: sepa-
                             rated by comma
<SRel>                   ::= srel (<Synonyms>, <Antonyms>, <Abbrevia-
                             tions>)
<Synonyms>               ::= [number] | list of Ids in [], if more: sepa-
                             rated by comma
<Antonyms>               ::= [number] |list of Ids in [], if more: sepa-
                             rated by comma
<Abbreviations>          ::= ['<string>'] | list of strings in [], if
                             more: separated by comma
<MorphoType>             ::= morpho, <Annotation>, <BasicEntry>, <Mor-
                             phoId>.
<BasicEntry>             ::= number , the Id of the corresponding basic
                             form

<Incomplete>             ::= incomplete, <Annotation>
<Annotation>             ::= any string, if none = empty

<MorphoRules>            ::= morpho_<x>(<MorphoId>, <PartOfSpeech>, <In-
                             flexion Class> ).
<MorphoId>               ::= number of morpho entry, unique for each lan-
                             guage, attached by the HyperLAT system
<PartOfSpeech>           ::= {m, f, n, v, adj, adv, num, prep, pron, conj,
                             art} (so far for German and Bulgarian)
<InflexionClass>         ::= <class symbol> of a grammar list
<Specification>          ::= NounSpecifics | VerbSpecifics | AdjeSpecifics
                             | PronSpecifics | PrepSpecifics
```

```
<NounSpecs>              ::= <SingPlr> <Szet> <Genitive><Dative> one or
                             more of the Specs with squared brackets,
                             separated by commas (so far only for German)
<SingPlr>                ::= {sg, pl} for singularetantum, pluraletantum
                             (so far only for German)
<Szet>                   ::= {ss} for the change of "ß" to "ss", when end-
                             ings follow (so far only for German)
<Genitive>               ::= {es, ses} when in genitive only -es is used
                             or the use is unstable (so far only for Ger-
                             man)
<Dative>                 ::= {e, oe} when in dative -e is used or the use
                             is unstable (so far only for German) (Atten-
                             tion: letter "o" not zero)

<VerbSpecs>              ::= <SepPrefix> <VerbType> <VerbChar>
<SepPrefix>              ::= '<string>' , a separable first part of the
                             verb (so far only for German)
<VerbType>               ::= {aux, mod} Default: full verb
<VerbChar>               ::= {refl, intr, impers} one or more of these
                             characteristics, l+trans+personal

<AdjSpecs>               ::= not yet defined

<PronSpecs>              ::= <pronType> ,<PronStructure>
<pronType>               ::= {pers, refl, poss, demo, rel, inter, indef}
                             one or more of these types
<PronStructure>          :: = the PoS of the second component in a complex
                             pronoun

<PrepSpecs>              ::='<string>', the first part of a compound
                             preposition

<StructRules>            ::= struct_<x>(<StructId>, <StructClass>,
                             <CatChain>).
<StructId>               ::= number of struct entry, unique for each lan-
                             guage, attached by the HyperLAT system
<StructClass>            ::= <class symbol> for the following <CatChain>
<CatChain>               ::= categorial chain, syntactic information for
                             the complex entry (>one word) (comment: The
                             grammatical head must be capitalized)

<SyntaxRules>            ::= synt_<x> (<QuotingEntry>, [<Information>] ).
                             (comment: The Id is the Id of the entry de-
                             scribed by this )
<Information>            ::= symbols, if more than one: separated by comma

<PhrasesRules>           ::= phrase_<x>(<PhraseId>, <Phrase>, <QuotingEn-
                             try>).
<PhraseId>               ::= number of phrase entry, unique for each lan-
                             guage, attached by the HyperLAT system
<Phrase>                 ::= '<text>'
<SemRules>               ::= sem_<x> (<QuotingEntry>, [<Informa-
                             tion>]).(comment: The Id is the Id of the
                             entry described by this )
<Information>            ::= symbols or features, if more than one: sepa-
                             rated by comma

<Trans>                  ::= trans_<x>_<y>(<X-Ids>,<Y-Ids>, <Annotation>)
<X-Ids>                  ::= [<QuotingEntry>], list of Ids of the source
                             entries of the TransEntry, if more than one:
                             separated by comma
<Y-Ids>                  ::= [<QuotingEntry>], list of Ids of the target
                             entries of the TransEntry, if more than one:
                             separated by comma
```

```
<Text>                    ::= text_<x> (<QuotingEntry>, <ExplanationText>)
                              with the Id of a non-term word
<ExplanationText>)        ::= 'any text'
```

Terminals are given in italics

18 Examples

18.1 Main Entry of a Term:

entry_g(327, ['kritischer', 'Moment'], term, [], lrel([], [325, 328]), none, 1, srel([], [], [])).

entry	Prefix for lexical entries
_g(	language indicator
327,	entry Id, unique number inside language "x"
[kritischer, 'Moment'],	Body of the entry
term,	Type, out of {word, term, morpho, incomplete}
[],	Specifics: here a non-instantiated Prolog variable
lrel	Prefix for Parts: "contained-in" and "consisting-of"
([], [325, 328]),	Entry 327 is lexically no part of any other entry [], and consists of 325 "kritisher" and 328 "Moment"
none,	no MophoId, complex entries have SyntaxId
1,	SyntaxId
srel	Prefix for semantic relations
([], [], [])	no Synonyms [], no Antonyms [], no Abbreviations
[]	
).	End of Entry

Comparison to the old lexicon format:

Old:	lex_entry_g(327, [kritischer, 'Moment'], term, _1, crossref([], [325, 328]), none, 1, sem([], [], none)).

Changes:
= both "kritischer" and "Moment" are quoted (as a contrast to the old format)
= the prefix part "lex_" is omited for brevity
= lrel (replaces the "crossref", stands for "lexical relations")
= srel (replaces the "sem", stands for "semantic relations").
= Abbreviations became a list instead of a single term

18.2 Main Entry of a Word:

lex_entry_g(288, [schwer], word, _1, lrel([1045], []), 18, none, srel([], [], none)).

```
entry                          Prefix for lexical entries
_g(                            language indicator
288,                           entry Id, unique number inside language "x"
['schwer'],                    Body of the entry
word,                          Type, out of {word, term, morpho, incomplete}
_1,                            Annotation: here a non-instantiated Prolog
                               variable
lrel                           Prefix for Parts: "contained-in" and "consist-
                               ing-of"
([], []),                      Entry 288 is lexically no part of entry 1045
                               [1045], and
                               consists of no other German entries []
18,                            MophoId
none,                          no SyntaxId, simple entries have only MorphoId
srel                           Prefix for semantic relations
([], [], [])                  no Synonyms [], no Antonyms [], no Abbrevia-
                               tions []
).                             End of Entry
```

Comparison to the old lexicon format:

```
Old:        lex_entry_g(288, [schwer], word, _1, crossref([1045], []), 18,
            none, sem([], [], none)).
```

Changes:

- = "schwer" is quoted (as a contrast to the old format)
- = the prefix part "lex_" is omited for brevity
- = lrel (replaces the "crossref" stands for "lexical relations")
- = srel (replaces the "sem" stands for "semantic relations").
- = Abbreviations became a list instead of a single term

18.3Entry for short reference:

entry_g(329, ['unkritisch'], incomplete, _1).

```
entry                          Prefix for lexical entries
_g(                            language indicator
329,                           entry Id, unique number inside language "x"
['unkritish'],                 Body of the entry
incomplete                     Type, out of {word, term, morpho, incomplete}
_1                             Annotation: here a non-instantiated Prolog
                               variable
).                             End of Entry
```

Comparison to the old lexicon format:

```
Old:        lex_entry_g(329, [unkritisch], incomplete, _1).
```

Changes:

- = the body "unkritish" is quoted (in contrast to the old format)
- = the prefix part "lex_" is omited for brevity.

18.4Entry for Inflected Forms:

entry_g(190, ['diesen'], morpho, _1, 188, 26).

```
entry                          Prefix for lexical entries
_g(                            language indicator
190,                           entry Id, unique number inside language "x"
['diesen'],                    Body of the entry
```

```
morpho                          Type, out of {word, term, morpho, incomplete}
_1,                             Annotation: here a non-instantiated Prolog
                                variable
188,                            Id of the entry containing the citation form of
                                "diesen";
                                entry 188 is "dieser" in the German lexicon
26                              MorphoId of 188 "dieser"
).                               End of Entry
```

Comparison to the old lexicon format:

```
Old:        lex_entry_g(190, [diesen], morpho, _1, 188, 26).
```

Changes:

- = the body "diesen" is quoted (as a contrast to the old format)
- = the prefix part "lex_" is omited for brevity

18.5 Entry for Morphological Rules:

morpho_g(66, n, n30).

```
morpho                          Prefix for morphological rules
_g(                             language indicator
66,                             Morpho Id, Id of the Morpho class
n,                              Part of Speech
[n30]                           Morpho class
).                              End of Entry
```

Comparison to the old lexicon format:

```
Old:        lex_morpho_g(66, n, [n30]).
```

Changes:

- = the prefix part "lex_" is omited for brevity

18.6 Entry for Structural Rules:

struct_g(1, 'np1', ['adj', 'N']).

```
syntax                          Prefix for structural entries
_g(                             language indicator
1,                              StructId, Id of the constituen structure
'np1',                          Categorial class, abbreviation
['adj', 'N']                    Syntax structure of a complex entry (e.g. "kri-
                                tischer Moment"), the head of the phrase
                                is capitalized, any element inside the struc-
                                ture is quoted
).                              End of Entry
```

Comparison to the old lexicon format:

```
Old:        lex_syntax_g(1, np1, [adj, 'N']).
```

Changes:

- = "struct" is the new entry prefix instead of "lex_syntax"
- = each string is quoted, not only the capitalized constituent element

18.7 Entry for Explanations of Words:

text_g(54, 'sich ohne Hilfsmittel auf der Oberfläche einer Flüssigkeit bewegen').

text	Explanation Entry
_g(	language indicator
54,	Quoted Id of an entry (Type=word), i.e. the entry to which the text belongs, whose meaning is described in the 2nd argument by the text string 'sich ... bewegen''sich ohne Hilfsmittel auf der Oberfläche einer Flüssigkeit bewegen' textual definition of the meaning of the German entry 54
).	End of Entry

Comparison to the old lexicon format:
Old: lex_text_g(54, 'sich ohne Hilfsmittel auf der Oberfläche einer Flüssigkeit bewegen').

Changes:
= the prefix part "lex_" is omited for brevity

18.8Entry for Translations:

trans_b_g(27, [trans(nul, [])], _1).

trans	Prefix for Translations
_b(	Source language indicator
_g(	Target language indicator
27,	Id of source language entry (_b 27) whose translation correspondences are defined by the next argument
[35, 64],	Ids of translation equivalents in German of the Bulgarian entry 27
_1	Annotation: here a non-instantiated Prolog variable
).	End of Entry

Comparison to the old lexicon format:
Old: trans_b_g(27, [trans(nul, [])], _1).

Comments and Changes:
Although the translation enries in the DB-MAT lexicon are not consistent at the moment, the idea there was to allow for a definition of translations when the corresponding entry is missing (e.g. to keep some translations as texts inside the "trans_x_y" = clause). Another idea was to distinguish the positions inside the source-target pairs, so to be able to say: "A" is translation of "B" and "word 1 of A" is the translation of "word 3 of B". But these ideas made the structure rather complicated.

That is why the current structure provides translation schema "one to many". To cover cases of "many to many", subsequent "trans_x_y" clauses must be included. Note that the clauses "trans_x_y" and "trans_y_x" provide enough space for definition of different translation complications when they are used flexibly.

18.9Entry for Syntax:

synt_r(72, [a, b, c, d, ...], _1).

synt	Prefix for Syntax
_r(	language indicator
72,	Id of a basic entry whose syntactic features are defined by the next argument
[a, b, c, d,...],	syntactic features for entry Romanian 72.
_1	Annotation: here a non-instantiated Prolog variable
).	End of Entry

Comparison to the old lexicon format:
This entry type did not exist in the DB-MAT lexicon. We assume that for the goals for parsing, much more syntactic information might be necessary for Romanian. The clause synt_r/3 allows for description of entry-centered syntactic information (e.g. feature matrices for unification). If some other syntactic information is needed, say for sentences, it may be organized as special part of the DBR-MAT lexicon with new clause-names (or special data fragments in the parser itself etc.).

18.10 Entry for Semantic Descriptions:

sem_g(127, [y, n, n]).

sem	Prefix for Semantics
_g(	language indicator
127,	Quoted Id of a basic entry whose semantic features are defined by the next argument
[y, n, n]	Values of semantic features of entry 127: for instance,animated="Yes", countable="No", mass-noun="No"
).	End of Entry

Comparison to the old lexicon format and comments:
This entry type did not exist in the DB-MAT lexicon. The list of semantic features can be different for different languages so that the interpretation and the encoding of this clause is open for multilingual design.

18.11 Entry for Phrases:

phrase_g(64, 'warmes Gericht', 963).

phrase	Prefix for phrasal entries
_g(	language indicator
64,	Id of phrasal entries, unique number in each language
'warmes Gericht',	phrase (collocation or idiom) which is worth to be remembered as text related to a basic entry
963	Quoted Id of a basic entry (here "Gericht" whose meaning as "meal" is remembered in this way as opposed to ist legal/terminological meaning).
).	End of Entry

Comparison to the old lexicon format:
This entry type did not exist in the DB-MAT lexicon. Please note that storing the collocations and idioms as texts restriict their use to strings shown to the user only.

19 Indices

19.1 Highest Lexical Entry-Id: entry_index_g(341).

```
entry_index                 Prefix for the highest entry number
_g                          language indicator
(341).                      Highest Entry number
```

Comment: Note that several types of lexicon clauses are "hooked to" the number-space of the basic lexical entries: sem_x, synt_x, text_x, phrases_x. That is why their Ids are not bigger than the current value of entry_index_x(X). The other two types of lexicon clauses however have independent enumeration and need their own indices.

19.2 Highest Morpho Entry-Id: morpho_index_g(74).

```
morpho_index                Prefix for the highest morpho number
_g                          language indicator
(74).                       Highest morpho number
```

19.3 Highest Structural Entry-Id: struct_index_g(7).

```
struct_index                Prefix for the highest syntax number
_g                          language indicator
(7).                        Highest number of complex term syntax
```

20 File Hierarchy of the HyperLAT Files

HyperLAT System

Grammar Data
- Grammar Data_g
  - PoS_g
  - Term_g
  - ------
  - MascGrammar_g
  - FemiGrammar_g
  - NeutGrammar_g
  - VerbGrammar_g
  - AdjeGrammar_g
  - AdveGrammar_g
  - NumeGrammar_g
  - ArteGrammar_g

PrepGrammar_g
ConjGrammar_g
PronGrammar_g
(and/or others)
---
nounSpecifics_g
VerbSpecifics_g
AdjeSpecifics_g
AdveSpecifics_g
NumeSpecifics_g
ArteSpecifics_g
PrepSpecifics_g
ConjSpecifics_g
PronSpecifics_g
(and/or others)

Grammar Data_b
PoS_b
Term_b

MascGrammar_b
FemiGrammar_b
neutGrammar_b
VerbGrammar_b
AdjeGrammar_b
AdveGrammar_b
NumeGrammar_b
ArteGrammar_b
PrepGrammar_b
ConjGrammar_b
PronGrammar_b
(and/or others)
---
nounSpecifics_b
VerbSpecifics_b
AdjeSpecifics_b
AdveSpecifics_b
NumeSpecifics_b
ArteSpecifics_b
PrepSpecifics_b
ConjSpecifics_b
PronSpecifics_b
(and/or others)

Grammar Data_r
PoS_r
Term_r

MascGrammar_r
FemiGrammar_r
neutGrammar_r
VerbGrammar_r
AdjeGrammar_r
AdveGrammar_r
NumeGrammar_r
ArteGrammar_r
PrepGrammar_r
ConjGrammar_r
PronGrammar_r
(and/or others)
---
nounSpecifics_r

VerbSpecifics_r
AdjeSpecifics_r
AdveSpecifics_r
NumeSpecifics_r
ArteSpecifics_r
PrepSpecifics_r
ConjSpecifics_r
PronSpecifics_r
(and/or others)

Grammar Data_p
PoS_p
Term_p

MascGrammar_p
FemiGrammar_p
neutGrammar_p
VerbGrammar_p
AdjeGrammar_p
AdveGrammar_p
NumeGrammar_p
ArteGrammar_p
PrepGrammar_p
ConjGrammar_p
PronGrammar_p
(and/or others)
---
nounSpecifics_p
VerbSpecifics_p
AdjeSpecifics_p
AdveSpecifics_p
NumeSpecifics_p
ArteSpecifics_p
PrepSpecifics_p
ConjSpecifics_p
PronSpecifics_p
(and/or others)

Lexicons
Lexicons-g
Lexicons-b
Lexicons-r
Lexicons-p

Word Lists
Word List_g
Word List_b
Word List_r
Word List_p

21 References

1. Walther v.Hahn / Galja Angelova: Providing Factual Information in MAT. In: Proceedings of "MT - 10 Years on". Cranfield 1994

2. Walther v.Hahn / Galja Angelova: System Architecture and Some System-Specific Components in Knowledge Based MAT. Project DB-MAT, Technical report 1-94, University of Hamburg, Computer Science Department, July 1994.

3. Walther v.Hahn. Analysis of Translation Errors in German-Bulgarian Translation of Domain-Specific texts. Project DB-MAT, Technical report 2-94, University of Hamburg, Computer Science Department, Sept. 1994. (zusammen mit Galia Angelova)

4. Walther v.Hahn: Die Verbindung lexikalischer und terminologischer Definition. Lösungen in einem Übersetzungs-Unterstützungssystem. In: In: Knobloch, C. und B. Schaeder (Hrsg), Nomination - fachsprachlich und gemeinsprachlich. Opladen 1996. S. 187 - 214

5. Walther v.Hahn: Combining Terminology, Lexical Semantics and Knowledge Representation in Machine Aided Translation. In: Christian Galinski & Klaus-Dirk Schmitz (eds.) TKE'96, Terminology and Knowledge Engineering. Frankfurt1996. S. 304 – 314 (zusammen mit Galja Angelova)

6. Walther v.Hahn: A Method to Avoid Disparate Answers from a Language Knowledge Base. Project “German.Bulgarian-Romanian Machine Aided Translation”, Working Paper 1-96 Februar 1996

7. Walther v.Hahn: Technical Language and its Translation. Project “German-Bulgarian-Romanian Machine Aided Translation”, Technical Report HH-1-97, May 1997

8. Walther v.Hahn: Handling Multilingual Technical Terms in a Knowledge Based System . In Gyde Hansen (Hrsg) LSP Texts and the Process of Translation. Kopenhagen 1998. S. 31 - 57