German-English-Romanian Lexicons (G.E.R.L.)

Report 2.0
September 2005

Monica Roxana Gavrila
(University of Hamburg, NATS Department)

Introduction

G.E.R.L. project consists of German, English, and Romanian Lexicons and of the bilingual connections between them. The lexicons should be encoded in such a way that can be used in several application types (Machine Translation, etc). The first main purpose of these lexicons is to be used by practicum students. One of the requirements of these lexicons is to have a standard structure, so that they fit in the actual stage of Natural Language Processing (NLP) applications. The lexicons contain only lexemes.

The idea from the beginning was to create G.E.R.L. having the MILE structure (Mono-Mile structures connected between them). Analyzing the MILE structure, I could not find information on Morphological Unit (MU) (the main material studied was the MILE Report (Deliverable D2.2-D3.2)). I also asked persons that worked with/at this structure about the MU. The general answer was that there is no real Morphological Unit, and that they worked more with PAROLE/SIMPLE structure, that is compatible with MILE structure.

After obtaining these answers, it has been decided to follow the PAROLE/SIMPLE structure. Another reason for choosing this is that there already exist German and English lexicons (partial lexicon entries can be found at http://www.ub.es/gilcub/SIMPLE/simple.html).

Specification

According to the starting purpose of the lexica, G.E.R.L. should contain the following information:

- **Morphology:**

  Part of speech
  - Noun: type, gender, number, case, morphological segmentation (suffixes, prefixes)
  - Verb: type, mode, tense, voice, number, way of saying if it is with particle or not (German and English)
  - Pronoun: type, person, gender, number case
  - Adjectives: gender, number, case, degree
  - Article: gender, number, case, type
  - Adverb: type
  - Numeral: type
  - Preposition
  - Conjunction
  - Verb particle (English and German)
- **Syntax:**
  - Cases for prepositions
  - Main/subordinate sentences for conjunctions and verbs
  - Personal / not personal verbs
  - Transitive / intransitive verbs
  - Mass nouns: nouns with only singular, or plural, or uncountable
- **Semantics:**
  - Synonyms
  - Thematic roles for verbs
  - Collocations
  - Way of saying if a word is foreign or no

G.E.R.L. is thought to be a full-form lexicon.

In case of compound words, all the words in the compound one should be already in the dictionary. The part of speech for a compound word is the one of the MAIN word. In case of no possibility of connection between the languages it is said that is a lexical gap.

Because it is followed the PAROLE/SIMPLE structure (SGML encoded), the above structure can be easily changed by modifying the DTD.

Due to the Romanian language and of the lexicon specification, there were made changes in the initial PAROLE/SIMPLE DTD.

Information on the Romanian language can be found in the precedent G.E.R.L report, as well as on MULTEXT and MULTEXT-EAST. The Romanian language was studied for the BALRIC-LING project/papers, MULTEXT-EAST and BalkaNet.

WORDNET

WordNet (*http://www.globalwordnet.org/*) exists for several languages, including Romanian (BalkanNet), English (WordNet 2.0, EuroWordNet), and German (EuroWordNet).

As mentioned on the Princeton WordNet and EuroWordNet websites[1]: "WordNet® is an on-line lexical reference system [...]. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets." "The word-nets are linked to an Inter-Lingual-Index. Via this index, the languages are interconnected so that it is possible to go from the words in one language to similar words in any other language." In WordNet, the existing information and relations between synsets are not enough for the goal of the lexicon - e.g. more morphological information needed, more (technical) words to be introduced, etc.

The G.E.R.L. structure

This section is describing the G.E.R.L. structure. As being mentioned above, the starting point in creating the G.E.R.L. structure is the PAROLE/SIMPLE DTD (http://gilc.ub.es/DTD-ALL/index.html). The original structure was simplified according to the specification needed and several features were added so that the problems due to the Romanian language are solved.

The G.E.R.L. structure is composed of a morphological layer, a syntactic layer, a semantic layer and a multilingual one (as SIMPLE/PAROLE structure). The first 3 layers have main units. Each unit has a unique id (attribute).

Morphological layer:

The main unit of this layer is the Morphological Unit(MU). From the original 4 types of MUs, there were kept 3:

1. Simple MU (MuS): for simple words entries
2. Compound MU (MuC): for compound words
3. Affix MU (MuAff): for affixes (this will help describing which noun has affixes, in the Derivation tag: Derivation / RDeriv)

Part of speech is given by the attribute gramcat; most of the types by the attribute gramsubcat. The word is contained in a new introduced tag in the MuS and MuC: Entry. Morphological features are given by the inp attribute of MUs that makes the connection to GInP. In GInP we have CombMFCif with attribute combmf, where morphological features are specified. Also in GInP can be specified number problems for nouns (uncountable, etc).

Syntactic layer:

The main unit is the Syntactic Unit (SynU). The cases for prepositions are described here: SynU / Description / Construct / SyntFeatureClosed / case. In the same way, with small modification of the existing DTD can be specified the verb main / subordinate clause problem.

Semantic layer:

The main unit is the Semantic Unit (SemU). In this layer synonyms are specified (as synonym relation between SemUs: SemU / RWeightValSemU / semR -> RsemU). In this

[1]http://wordnet.princeton.edu/w3wn.html (Princeton WordNet), http://www.illc.uva.nl/EuroWordNet/ (EuroWordNet).

layer thematic roles for verbs are described: SemU / PredicativeRepresentation / Predicate / Argument / Semantic role.

To specify collocations a new tag Collocation was introduced in the semantic part. It is in such a way built so that bilingual connections can be easily realized.

If a word is foreign or not it is specified in the MUs (it is somehow independent of the syntactic / semantic behavior of the word). This is a difference comparing to the original structure.

There is the possibility to link MUs to SynUs (the synulist attribute in MUs) and to SemUs.

Multilingual layer:

The existing multilingual layer was modified, due to the Romanian language and due to the specification of the lexica. Due to the Romanian language, it should exist the possibility to connect Mu. In the existing DTD the connection is at SynU and SemU (concepts, etc - but these connections are not taken into consideration for these lexica). In this project connection are made at the following levels : MUs and Collocation. Also at this level is mentioned the lexical gap problem (CorrespGap) : is not a connection to a correspondent MU, but it is given the translation (went - ist gegangen).

Wheelchair vs. scaun cu rotile: wheelchair as MuC and there are links to the words wheel and chair. This way the translation is logical

The multilinguality connection is not always bi-directional (e.g. wheelchair -> scaun cu rotile). For words where both MUs are specified is bi-directional, else is only in one direction (it is given the translation).

The structure of the G.E.R.L. can be seen in Figure 1 and in the DTD.

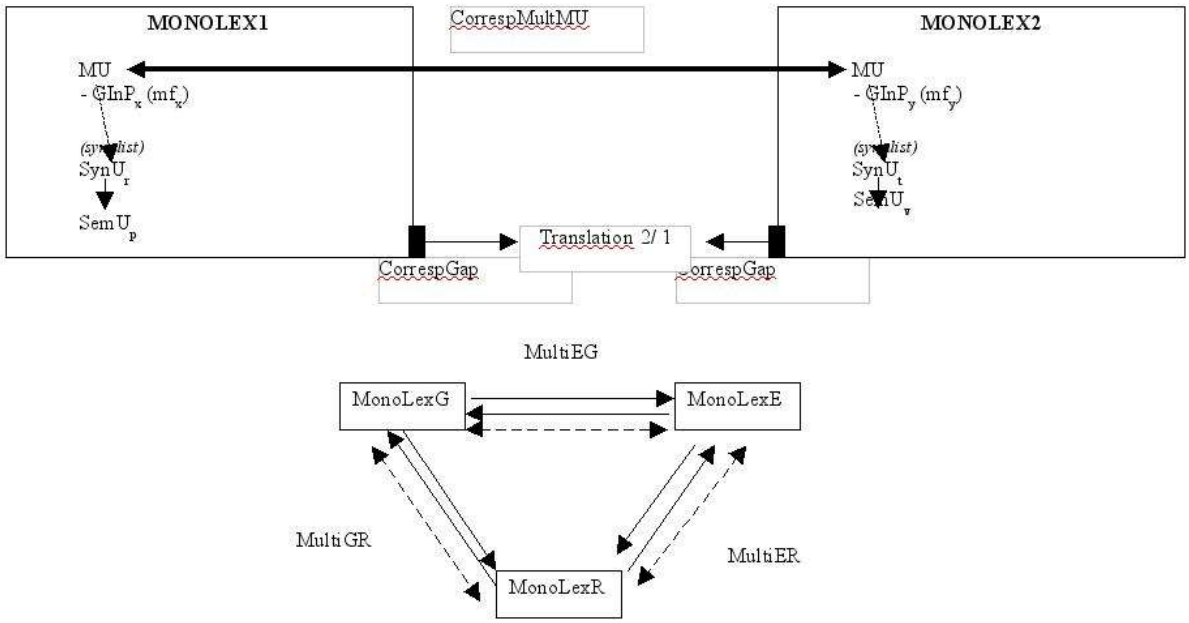


Figure 1. G.E.R.L Structure

Software Details

The software is implemented in Java (java version "1.5.0_04". It works also on java 1.4.). It was tested on Windows, Linux (Debian) and Mac, but on a very limited set of operations.
It uses jdom-1.0 – for managing the XML file

Downloads:
Java: http://java.sun.com/downloads/index.html
JDOM: www.jdom.org

The operations that should be done with the G.E.R.L. tool are:
- adding entries
- deleting entries
- viewing/updating entries
- viewing the lexicon file
- getting statistics about the lexicon file
- updating lexicon information
- updating collocations

At the moment the tool is not fully working. The extensions that have to be done are presented below:
1. Update MuS – At the moment should be done manually
2. Update MuC - At the moment should be done manually
3. Modify the tool so that it deals also with translations of the type 1:n (for all: MuS, MuC, Collocations) – IF NECESSARY. At the moment it deals with translation of the type 1:1.
4. Adding operations should be extended to all PoS - It is working only for Verb at full capabilities (This means including multilingual and semantic information).
5. When deleting an entry, it should delete also collocations or translations connected to this word (in a logical way!) - At the moment should be done manually
6. The GUI should be more user friendly. For this JTextFields should be transformed in JLists – where possible
7. The tool might contain bugs. It should be tested.

Lexion Details

Number of entries in the lexicons:
German: 441 (Simple entries: 440 MuS, Compund entries: 1 MuC)
Romanian: 0
English: 0
Multilingual connections:0

Part of speech Information

PoS
Foreign word
Morphology
Syntax
Semantics
Multilinguality

1. Noun
   Foreign
   Type, Gender, Number, Case, Article, Derivation
   -

Synonyms, Collocations
Multilingual information

2. Verb
   Foreign
   Type, Mode, Number,Tense,Voice,Person, Transitivity, Type (im/personal), Particle
   Sentence
   Synonyms, Collocations, Thematic Roles
   Multilingual information
3. Pronoun
   Foreign
   Type, Gender, Number, Case,Person
   -
   Synonyms, Collocations
   Multilingual information
4. Adjective
   Foreign
   Type, Gender, Number, Case,Degree, of, Article
   -
   Synonyms, Collocations
   Multilingual information
   It should be necessary adding type – at least in Romanian!!!!! To modify the DTD if necessary! In this case, being a full form lexicon, it is not really necessary.
5. Article
   Type, Gender, Number, Case
   -
   -
   Multilingual information
6. Adverb
   Foreign
   Degree
   -
   Synonyms, Collocations
   Multilingual information
7. Numeral
   Foreign
   Type, Gender, Case
   -
   Synonyms, Collocations
   Multilingual information
8. Preposition
   Foreign
   -
   Case restrictions
   -
   Multilingual information
9. Conjunction
   Foreign
   Type
   -
   -
   Multilingual information
10. Verb Particle
    Foreign

-
-
-
Multilingual information

11. Affix
-
Type
-
-
-

Example of entry in the English Lexicon with connection to the Romanian lexicon

```
<?xml version="1.0" encoding="UTF-8"?>
<LesParole>
 <Parole>
  <ParoleMorpho>
   <MuS gramcat="Verb" subgramcat="main" id="Ver_0001" synulist="V---not known--" semulist="EMPTY" foreign="NO">
    <Entry>test</Entry>
    <Gmu inp="V-infinitive-WITHOUT-WITHOUT-WITHOUT-WITHOUT-WITHOUT-WITHOUT-No" />
   </MuS>
   <GInp id="V-infinitive-WITHOUT-WITHOUT-WITHOUT-WITHOUT-WITHOUT-WITHOUT-No">
    <CombMFCif combMF="V_infinitive_WITHOUT_WITHOUT_WITHOUT_WITHOUT_WITHOUT_WITHOUT_No" />
   </GInp>
   <CombMF id="V_infinitive_WITHOUT_WITHOUT_WITHOUT_WITHOUT_WITHOUT_WITHOUT_No" mood="infinitive" tense="WITHOUT" voice="WITHOUT" number="WITHOUT" person="WITHOUT" transitivity="WITHOUT" typepers="WITHOUT" hasparticle="No" />
  </ParoleMorpho>
  <ParoleSyntaxe>
   <SynU id="EMPTY" comment="no syntactical information" example="" description="EMPTY" />
   <Description id="EMPTY" comment="" example="" />
   <SynU id="V---not known--" comment="V with restriction on --not known--" example="no example" description="V_--not known--" />
   <Description id="V_--not known--" comment="no comment" example="no example" representativemu="no example" construction="V/--not known--" />
   <Construction id="V/--not known--" comment="no comment" example="no example">
    <SyntFeatureClosed featurename="FOLLOWEDBY" value="--not known--" />
   </Construction>
  </ParoleSyntaxe>
  <ParoleSemant>
```

```
    <SemU id="EMPTY" comment="no semantic information" example="" collocationlist=""
/>
    <RSemU id="SYN" comment="synonymy relation" sstype="SYNONYMY" />
    <SemanticRole id="SR_agent" example="" comment="" name="agent" />
    <SemanticRole id="SR_patient" example="" comment="" name="patient" />
    <SemanticRole id="SR_experiencer" example="" comment="" name="experiencer" />
    <SemanticRole id="SR_theme" example="" comment="" name="theme" />
    <SemanticRole id="SR_location" example="" comment="" name="location" />
    <SemanticRole id="SR_instrument" example="" comment="" name="instrument" />
    <SemanticRole id="SR_source" example="" comment="" name="source" />
    <SemanticRole id="SR_goal" example="" comment="" name="goal" />
   </ParoleSemant>
  </Parole>
  <ParoleMultilingue langue1="English" langue2="German" />
  <ParoleMultilingue langue1="English" langue2="Romanian">
   <CorrespMultMU id="CorrespMU_1" commentaire="" mulangue1="Ver_0001"
mulangue2="Ver_0001" />
  </ParoleMultilingue>
</LesParole>
```

Example of the empty lexicon file (Romanian)

```
<?xml version="1.0" encoding="UTF-8"?>
<LesParole>
  <Parole>
   <ParoleMorpho />
   <ParoleSyntaxe>
    <SynU id="EMPTY" comment="no syntactical information" example=""
description="EMPTY" />
    <Description id="EMPTY" comment="" example="" />
   </ParoleSyntaxe>
   <ParoleSemant>
    <SemU id="EMPTY" comment="no semantic information" example="" collocationlist=""
/>
    <RSemU id="SYN" comment="synonymy relation" sstype="SYNONYMY" />
    <SemanticRole id="SR_agent" example="" comment="" name="agent" />
    <SemanticRole id="SR_patient" example="" comment="" name="patient" />
    <SemanticRole id="SR_experiencer" example="" comment="" name="experiencer" />
    <SemanticRole id="SR_theme" example="" comment="" name="theme" />
    <SemanticRole id="SR_location" example="" comment="" name="location" />
    <SemanticRole id="SR_instrument" example="" comment="" name="instrument" />
    <SemanticRole id="SR_source" example="" comment="" name="source" />
    <SemanticRole id="SR_goal" example="" comment="" name="goal" />
    <SemanticRole id="SR_no" example="" comment="" name="no semantic role" />
   </ParoleSemant>
```

```
 </Parole>
 <ParoleMultilingue langue1="Romanian" langue2="English" />
 <ParoleMultilingue langue1="Romanaian" langue2="German" />
</LesParole>
```

The G.E.R.L. DTD

Observation: The first G.E.R.L. DTD was a little bit different.

```
<!-- The original dtd was simplified to the needs of GERL - this means that several attributes
and tags were eliminated -->
<!-- Also there were added some tags in order to make the correspondances between
Romanian and other languages -->
<!-- Modified ? in " " at Parole-->

<!DOCTYPE LesParole [
<!ELEMENT LesParole - O ( Parole+ , ParoleMultilingue+)

<!ELEMENT Parole - O
        ( ParoleMorpho, ParoleSyntaxe, ParoleSemant)>
<!ATTLIST Parole
        lexiconname         CDATA                   #REQUIRED
        language            CDATA                   #REQUIRED
        version             CDATA                   #IMPLIED
        creationdate1       CDATA                   #IMPLIED
        modificationdate    CDATA                   #IMPLIED
        copyright           CDATA                   #IMPLIED>

<!-- ******************************************************** -->
<!-- *****         MORPHOLOGICAL INFORMATION         ***** -->
<!-- *****              ParoleMorpho              ***** -->
<!-- ******************************************************** -->
<!ELEMENT ParoleMorpho - O
                        ((MuS|MuC|MuAff)* &
                        GInP* &
                        CombMF*)>

<!-- ************************************************  -->
<!-- *******  DEFINITION OF MORPHOLOGICAL UNITS   ******  -->
<!-- ************************************************  -->
<!-- covered noun, verb, pronoun, adjective, article, adverbs, numerals, prepositions,
conjuctions, particles-->
<!--some of the types (noun, verb, person, numeral), main/subord for the conjuctions-->
```

```
<!ELEMENT MuS  - O (Entry,Gmu+ & Derivation*)>
<!--The ADPOSITION was tranformed in PREPOSITION and VERBPARTICLE-->
<!ATTLIST MuS
        id              ID                                  #REQUIRED
        gramcat                 (WITHOUT|NOUN|VERB|
                                ADJECTIVE|PRONOUN|
                                ADVERB|PREPOSITION|VERBPARTICLE|
                                CONJUNCTION|NUMERAL|
                                ARTICLE)            WITHOUT
        gramsubcat      (WITHOUT|PROPER|
                                COMMON|MAIN|AUX|MODAL|
                                COPULA|
                                POSSESSIVE|DEMONSTRATIVE|
                                INTERROGATIVE|RELATIVE|RECIPROCAL
                                REFLEXIVE|PERSONAL|UNDEFINED|NEGATIVE
                                COORDINATIVE|SUBORDINATIVE|
                                CARDINAL|ORDINAL|
                                FRACTIONAL|REPETATIVE|MULTIPLICATIVE|
VARIATIVE|
                                DEFINITE|INDEFINITE|OTHER)  WITHOUT
        synulist        IDREFS                              #IMPLIED
        foreign                 (YES|NO|NOSPEC)                 NOSPEC
        semulist        IDREFS                              #IMPLIED>
<!--added semulist as attribute. Connection Mu-SemU, no SynU-SemU
-->

<!ELEMENT MuC  - O (Entry, RCompos+)>
<!--The ADPOSITION was tranformed in PREPOSITION and VERBPARTICLE-->
<!ATTLIST MuC
        id              ID                                  #REQUIRED
        gramcat                 (WITHOUT|NOUN|VERB|
                                ADJECTIVE|PRONOUN|
                                ADVERB|PREPOSITION|VERBPARTICLE|
                                CONJUNCTION|NUMERAL|
                                ARTICLE)            WITHOUT
        gramsubcat      (WITHOUT|PROPER|
                                COMMON|MAIN|AUX|MODAL|
                                COPULA|
                                POSSESSIVE|DEMONSTRATIVE|
                                INTERROGATIVE|RELATIVE|RECIPROCAL
                                REFLEXIVE|PERSONAL|UNDEFINED|NEGATIVE
                                COORDINATIVE|SUBORDINATIVE|
                                CARDINAL|ORDINAL|
```

```
                                FRACTIONAL|REPETATIVE|MULTIPLICATIVE|
VARIATIVE|
                                DEFINITE|INDEFINITE|OTHER)  WITHOUT
        synulist        IDREFS                              #IMPLIED
        foreign                 (YES|NO|NOSPEC)                 NOSPEC
        mainword        IDREF                               #IMPLIED
        semulist        IDREFS                              #IMPLIED>
<!--added semulist as attribute. Connection Mu-SemU, no SynU-SemU
        A Compound Morphological Unit has no Gmu of its own:
        these graphic forms are deduced from the Units
        which make up the Compound Unit.
        Each Component that participates in the MuC is indicated
        by an RCompos relationship.
        A MuC consists at least of 2 Rcompos (which the DTD does not show)
Foreign attribute is added in both cases (MuS, MuC) -->

<!ELEMENT MuAff  - O (Entry)>
<!ATTLIST MuAff
        id                      ID
#REQUIRED
        typaff                  (WITHOUT|PREFIX|
                                        SUFFIX|BASE)
        WITHOUT>
<!--    The attribute, typaff records the type of
        a Morphological Affix Unit; in the case in which an affix
        may be typed only within its derivation context, this
        attribute will have the value, WITHOUTS.-->

<!ELEMENT Entry O O (#PCDATA)>
<!-- This element is added so that the search is done more rapid, espacially for machine
translations -->

<!-- ********************************************  -->
<!-- *******   MORPHOLOGICAL COMPOSITION **********  -->
<!-- ********************************************  -->

<!ELEMENT RCompos - O  EMPTY>
<!ATTLIST RCompos
        linearorder             NUMBER                      #REQUIRED
        gsepar                  (ATTAQUEG|HYPHEN|
                                        APOSTROPHE|SPACE|
                                        JOIN|HYPHENSPACE|
                                        HYPHENJOIN|HYPHENAPOSTROPHE|
```

```
                                        HYPHENSPACEJOIN|
                                        APOSTROPHEJOIN|
                                        SPACEJOIN)              ATTAQUEG
        mu                              IDREF
#REQUIRED>

<!--    The attribute 'mu' indicates a MuS/Cont/Agg/C (Um sub-
        classes) component which participates in the composition.
        The attribute 'linearorder' specifies
        the position of the component in the composition.
        The attributs 'g/psepar' ("graphic/phonemic seperators"),
        gives the list of possible separators which may appear
        before the component. -->

<!-- ************************************************ -->
<!-- *********    GRAPHICAL FORM ******* -->
<!-- ************************************************ -->
<!-- Modification: not used anymore -->

<!-- ************************************************ -->
<!-- *****  GRAPHIC SYSTEM OF INFLECTION *** -->
<!-- ************************************************ -->

<!ELEMENT Gmu - O (EMPTY)>
<!ATTLIST Gmu
        inp             IDREF                   #REQUIRED>


<!ELEMENT GInP - O (CombMFCif+)>
<!ATTLIST GInP
        id              ID              #REQUIRED
        comment                 CDATA                   #IMPLIED
        example                 CDATA                   #IMPLIED

<!ELEMENT CombMFCif - O EMPTY>
<!ATTLIST CombMFCif
        combmf          IDREF                   #REQUIRED>

<!--    A CombMFCif refers to a CombTM (Combination of Morphological Features) via the
   'combmf' feature.
        Added of degree - possible not necessary -->

<!ELEMENT CombMF - O EMPTY>
```

```
<!ATTLIST CombMF
        id              ID                                              #REQUIRED
        gender          (WITHOUT|MASCULINE|FEMININE|
                                NEUTER)         WITHOUTG
        number                  (WITHOUT|SINGULAR|PLURAL)
WITHOUT
        case            (WITHOUT|NOMINATIVE|GENITIVE|
                                DATIVE|ACCUSATIVE|VOCATIVE)
WITHOUT
        mood            (WITHOUT|INDICATIVE|IMPERATIVE|
                                INFINITIVE|PARTICIPLE|GERUND|
                                CONJUNCTIVE)                            WITHOUT
        tense           (WITHOUT|PRESENT|IMPERFECT|
                                PAST|PLUSQUEPARFAIT|PERFECTSIMPLE)
WITHOUT
        person          (WITHOUT|1|2|3)                                 WITHOUT
        reflexivity     (WITHOUT|RREFL|NOREFL)
WITHOUT
        degree          (WITHOUT|POSITIVE|
                                COMPARATIVE|SUPERLATIVE)                WITHOUT
        degreetype      (WITHOUT|SUPERIORITY|INFERIORITY|EQUALITY|ABSOLUTE)
        WITHOUT
        transitivity    (WITHOUT|TRANSITIVE|INTRANSITIVE)
WITHOUT
        typepers        (WITHOUT|PERSONAL|IMPERSONAL)
WITHOUT
        article         (WITHOUT|DEFINITE|INDEFINITE)
WITHOUT
        hasparticle     (YES|NO)                                        NO
>
<!--it is covered gender, number, case, mode, time, voice, person, degree, reflxivity,
transitivity, (im)personal-->

<!--    ********************************************** -->
<!--    *******  MORPHOLOGICAL DERIVATION    *********** -->
<!--    ********************************************** -->

<!ELEMENT Derivation - O (RDeriv+)>
<!ATTLIST Derivation
        comment         CDATA                   #IMPLIED>
<!--    The content token 'RDeriv' is used to record the
        different components of a derivation. Concurrent
        derivations are indicated by recording several Derivation
```

```
        elements on one derived Unit.-->

<!ELEMENT RDeriv - O EMPTY>
<!ATTLIST RDeriv
        linearorder             NUMBER                  #IMPLIED
        status                  (WITHOUT|PREFIX|
                                        SUFFIX|BASE)    WITHOUT
        mu                      IDREF                   #REQUIRED>
<!--    The field 'mu' indicates the component of the derivation.
        The attribute 'linearorder' indicates the range of the Mu in the
        derivation -->

<!-- ************************************************** -->
<!-- *****          SYNTACTIC INFORMATION         ***** -->
<!-- *****            PAROLESYNTAXE               ***** -->
<!-- ************************************************** -->

<!-- It is used for Conjuctions and Verbs. The other have the EMPTY SynU-->

<!ELEMENT ParoleSyntaxe - O (
                        SynU+ &
                        Description+ &
                        Construction*)>

<!-- ************************************************** -->
<!-- *****       SYNTACTIC UNIT, DESCRIPTION      ***** -->
<!-- ************************************************** -->

<!ELEMENT SynU - O EMPTY>
<!ATTLIST SynU
        id                      ID              #REQUIRED
        comment                 CDATA           #IMPLIED
        example                 CDATA           #IMPLIED
        description             IDREF           #REQUIRED>
<!--    SynU describes one syntactic behaviour of a Mu.
    One has to encode as many SynU as syntactic behaviours for a same Mu.
        - The attribute 'description' records the base description,
        - CorrespSynUSemU encodes the correspondence with the semantic level.
(CorrespSynUSemU*) from SynU deleted -->

<!ELEMENT Description - O EMPTY>
<!ATTLIST Description
        id                      ID              #REQUIRED
```

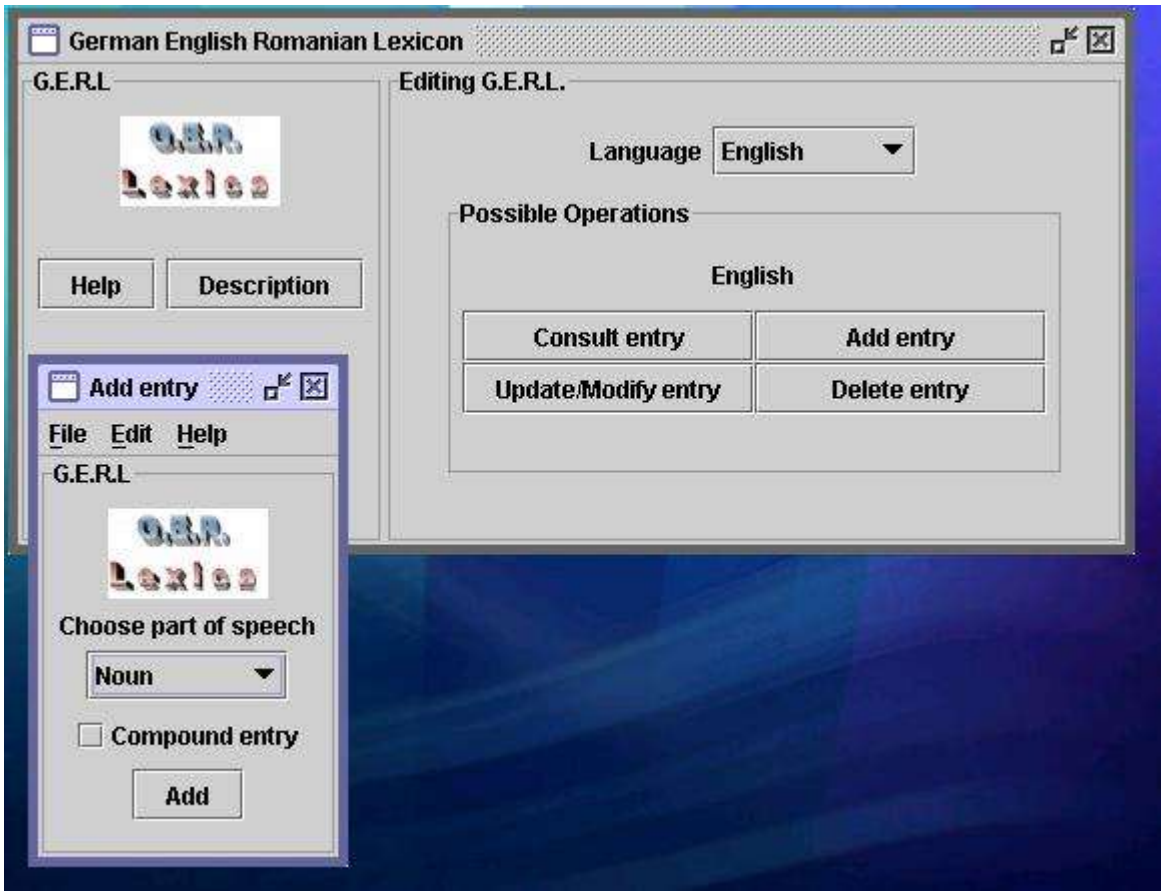
```
        comment                 CDATA                   #IMPLIED
        example                 CDATA                   #IMPLIED
        representativemu        CDATA                   #IMPLIED
        construction            IDREF                   #IMPLIED>
<!--    The attribute 'representativemu' records the id of MU,
        the attribute 'construction' records the id of the Construction -->

<!-- ********************************************************* -->
<!-- *****          CONSTRUCTION                 ***** -->
<!-- ********************************************************* -->

<!ELEMENT Construction - O (SyntFeatureClosed*)>
<!ATTLIST Construction
        id              ID                              #REQUIRED
        comment                 CDATA                           #IMPLIED
        example                 CDATA                           #IMPLIED>
<!-- A Construction describes the context or syntactic frame specific
    to the entry described. -->

<!-- ********************************************************  -->
<!-- *****            FEATURES              *****  -->
<!-- ********************************************************  -->

<!ELEMENT SyntFeatureClosed - O EMPTY>
<!ATTLIST SyntFeatureClosed
        featurename             (CASE|FOLLOWEDBY)
#REQUIRED
        value                   (MAIN|SUBORDINATE
                                        NOMINATIVE|GENITIVE|DATIVE|
                                        ACCUSATIVE|VOCATIVE)
#REQUIRED >


<!-- ********************************************************* -->
<!-- *****          SEMANTIC INFORMATION          ***** -->
<!-- *****            PAROLESEMANT              ***** -->
<!-- ********************************************************* -->

<!-- It is used for thematic roles  - VERBS, synonymy and collocations -->
<!-- Diffrent way of expressing collocations: new tag Collocation!!! -->

<!ELEMENT ParoleSemant - O (
```

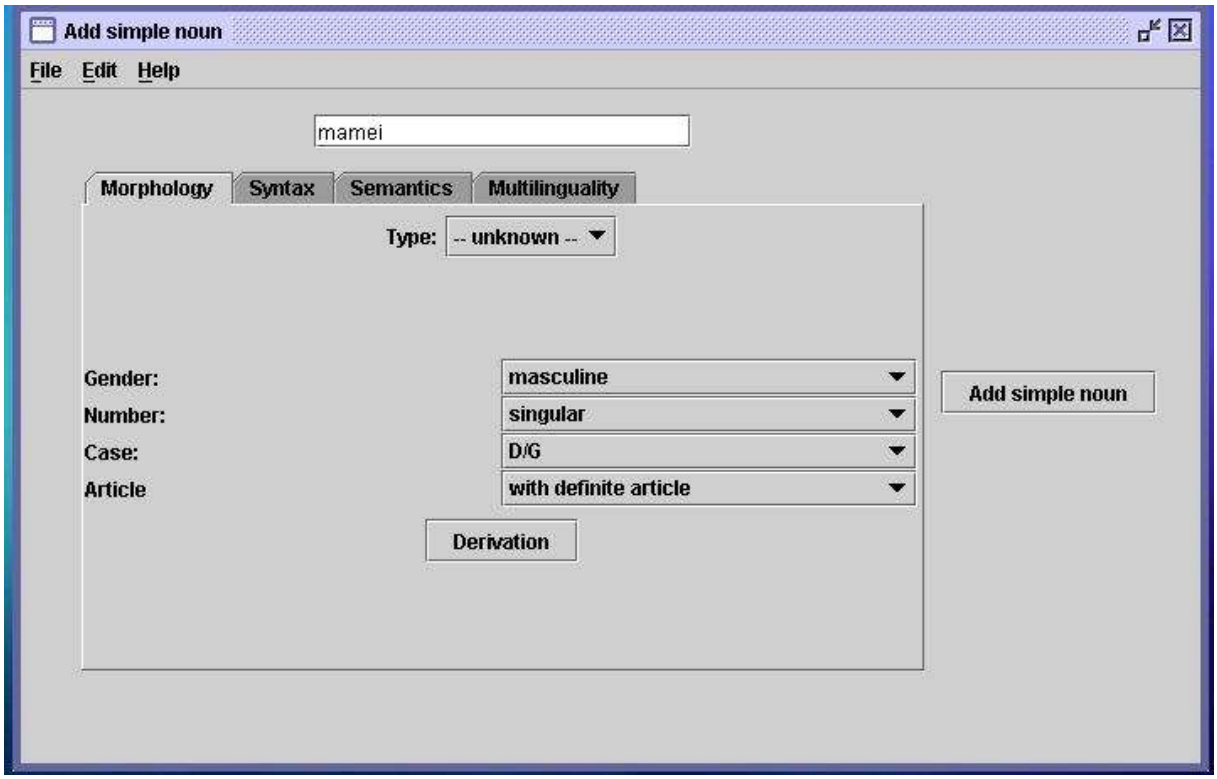
```
                    SemU+
                    & Predicate*
                    & Argument*
                    & SemanticRole*
                    & RSemU*
                    & Collocation*)>

<!ELEMENT SemU  - O (PredicativeRepresentation?,RWeightValSemU*) >
<!ATTLIST SemU
        id              ID              #REQUIRED
        example             CDATA               #IMPLIED
        comment             CDATA               #IMPLIED
        collocationlist     IDREFS          #IMPLIED>

<!ELEMENT PredicativeRepresentation - O EMPTY>
<!ATTLIST PredicativeRepresentation
        predicate       IDREF                   #REQUIRED>
<!-- Used for semantic/thematic roles -->

<!ELEMENT Predicate  - O EMPTY>
<!ATTLIST Predicate
        id              ID              #REQUIRED
        example             CDATA               #IMPLIED
        comment             CDATA               #IMPLIED
        argumentl       IDREFS          #REQUIRED>

<!ELEMENT Argument  - O EMPTY>
<!ATTLIST Argument
        id              ID              #REQUIRED
        example             CDATA               #IMPLIED
        comment             CDATA               #IMPLIED
        position1       (NO|BEFORE|AFTER)   NO
        position2       CDATA           #IMPLIED
        semanticrolel       IDREFS          #REQUIRED>
<!-- position1, position2 added for meking more clear -->

<!-- added already in the lexicon -->
<!ELEMENT SemanticRole  - O EMPTY>
<!ATTLIST SemanticRole
        id              ID              #REQUIRED
        example             CDATA               #IMPLIED
        comment             CDATA               #IMPLIED
        name            CDATA           #REQUIRED>
```

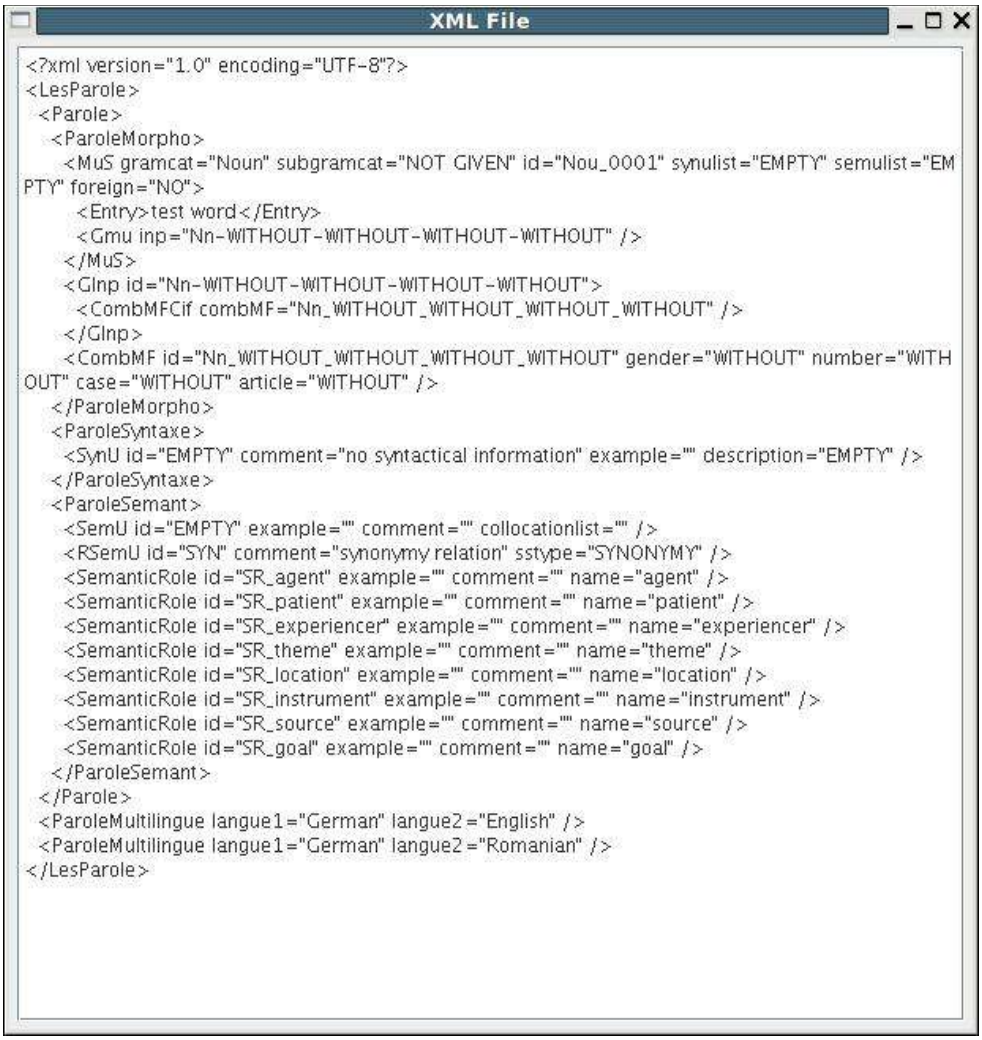
```
<!-- used for synonym-relation, modified target as a list because this way partial synonsmy is
also taken into consideration-->
<!ELEMENT RWeightValSemU - O EMPTY>
<!ATTLIST RWeightValSemU
        comment                 CDATA                   #IMPLIED
        targetlist      IDREFS                  #REQUIRED
        semr            IDREF                   #REQUIRED>

<!-- there is only one type of relation: synonymy
        added as such in lexicon - is the only one-->
<!ELEMENT RSemU - O EMPTY>
<!ATTLIST RSemU
        id              SYN             SYN
        comment                 CDATA                   #IMPLIED
        sstype          (SYNONYMY)              SYNONYMY>

<!-- new added tag so that entering collocations is easier -->
<!ELEMENT Collocation O O EMPTY>
<!ATTLIST Collocation
        id              ID                      #REQUIRED
        expression      CDATA                   #IMPLIED
        meaning                 CDATA                   #IMPLIED
        synonymMu       IDREF                   #IMPLIED>

<!-- ********************************************************** -->
<!-- *****           MULTILINGUALITY           ***** -->
<!-- ********************************************************** -->

<!ELEMENT ParoleMultilingue - O (CorrespMultColloc* & CorrespMultMU* &
CorrespGap*)>

<!ATTLIST ParoleMultilingue
        langue1                 CDATA                   #REQUIRED
        langue2                 CDATA                   #REQUIRED >

<!-- Added tags, almost totaly changed!!! -->

<!ELEMENT  CorrespMultColloc - O (Referent)>
<!ATTLIST CorrespMultColloc
        id                      ID              #REQUIRED
        commentaire             CDATA                   #IMPLIED
        colloclangue1           IDREF                   #REQUIRED>
```

```
<!ELEMENT Referent - O (EMPTY)>
<!ATTLIST Referent
        typereferent                (MU|TRANSLATION|COLLOCATION|NOTKNOWN)
                        NOTKNOWN
        referentref                 IDREF
        #IMPLIED
        translation                 CDATA
        #IMPLIED>
<!-- if there is no link to another one, here is the text
        typereferent was transformed from CDATA
-->


<!ELEMENT CorrespMultMU - O EMPTY>
<!ATTLIST CorrespMultMU
        id                          ID              #REQUIRED
        commentaire                 CDATA                   #IMPLIED
        mulangue1                   IDREF                   #REQUIRED
        mulangue2                   IDREFS                  #REQUIRED>
<!-- mulangue2 transformed from IDREF to IDREFS: translation 1:n. The interface is nor
supporting it
Translation word (MU) into collocations are given also in this tag. -->

<!ELEMENT CorrespGap - O EMPTY>
<!ATTLIST CorrespGap
        id                          ID              #REQUIRED
        commentaire                 CDATA                   #IMPLIED
        mu                          IDREF                   #REQUIRED
        translation                 CDATA                   #REQUIRED>

<!--    langueWithoutTranslation    CDATA                   #REQUIRED
        langueWithTranslation               CDATA                   #REQUIRED
Not used, because translation is only in one direction
-->
]>
```

Graphic Interface for G.E.R.L. Snapshots


German English Romanian Lexicon
G.E.R.L
Help
Description
Editing G.E.R.L.
Language
English
Possible Operations
English
Consult entry
Add entry
Update/Modify entry
Delete entry
Add entry
File
Edit
Help
G.E.R.L
Choose part of speech
Noun
Compound entry
Add


Add simple noun
File
Edit
Help
mamei
Morphology
Syntax
Semantics
Multilinguality
Type:
-- unknown --
Gender:
masculine
Number:
singular
Case:
D/G
Article
with definite article
Derivation
Add simple noun



References:

TEI website: www.tei-c.org.uk/P4X
MULTEXT: http://www.lpl.univ-aix.fr/projects/MULTEXT/
MULTEXT-East: http://nl.ijs.si/ME/
PAROLE/SIMPLE: http://www.ub.es/gilcub/SIMPLE/simple.html
BalkaNet: http://www.ceid.upatras.gr/Balkanet/
BALRIC-LING (for Romanian: RORIC-LING: http://phobos.cs.unibuc.ro/roric/)

G.E.R.L:
C. Vertan, W. von Hahn, M. Gavrila, "Designing a PAROLE/SIMPLE German-English-Romanian Lexicon", RANLP Workshop Bulgaria
G.E.R.L. Report