COMPUTATIONAL MODEL OF THE DICTIONARY ENTRY

Preliminary Report

ACQUILEX

ESPRIT BASIC RESEARCH ACTION No. 3030
Six Month Deliverable

Pisa, April 1990

ILC-ACQ-1-90

COMPUTATIONAL MODEL OF THE DICTIONARY ENTRY

Preliminary Report

Nicoletta Calzolari[1], Carol Peters[2], Adriana Roventini[3]

[1] Dipartimento di Linguistica, Università di Pisa, Italy
[2] Istituto di Elaborazione della Informazione, CNR, Pisa, Italy
[3] Istituto di Linguistica Computazionale, CNR, Pisa, Italy

ACQUILEX

ESPRIT BASIC RESEARCH ACTION No. 3030
Six Month Deliverable

Pisa, April 1990

ILC-ACQ-1-90

Contents

INTRODUCTION

The description of the computational model of the dictionary entry to be used within the ACQUILEX Project consists of two separate sections.
The first part, Section 1, presents a method which can be used to represent in a uniform way the content and structure of the entries of machine-readable dictionaries (MRDs), and contains an explicit standardized representation of the content of the different dictionaries being used within the Project.

In the last few years, the research community in the area of Computational Linguistics has become increasingly aware of the need to establish guidelines and standards for the representation of texts in machine-readable form. In fact, a number of international initiatives have been promoted with this explicit purpose, the most important being the Text Encoding Initiative (TEI 1989), cosponsored by the Association for Computational Linguistics (ACL), Association for Literary and Linguistic Computing (ALLC), Association for Computing in the Humanities (ACH) and by the European Community.

We feel that the definition of an explicit and uniform representation language for machine-readable dictionaries is essential for the following reasons:
- it allows the exchange of data in a common format;
- it makes possible uniform types of analyses of different dictionaries;
- it makes it possible to write parsers for different dictionaries on the basis of a common format;
- it makes it possible to preserve the source text (the dictionaries) for further applications;
- it makes it easier to standardize the contents (the values) of some of the fields;
- it facilitates the design of a common model of the Lexical Entry for the Project Database, into which all the existing dictionary representations must be mapped, on the basis of the representations of the existing data;
- it makes it possible to load the existing dictionaries into the common Project Lexical Database.

The methodology that we propose here will be applied before the end of the Project to other mono- and bilingual dictionaries, and revisions to the document are therefore to be expected.

The second part, Section 2, consists in a description of the common Project Lexical Database Entry. This will be improved and/or augmented during the different stages of development of the database itself, in accordance with the requirements of the Project. This description is more complex than that of Section 1, because it contains not only a description of the Entry following the same formalism used in Section 1 but also a brief specification of the database model which has been decided on by the Project.

The two sections together constitute our integrated proposal for the Computational Model of a Dictionary Entry.

Note

In the present document, we present a preliminary report on the computational model of the dictionary. The aim is to produce a working definition of a general representation of a lexical entry capable of handling not only the dictionaries actually in use in the Project but also other MRD sources which may be made available to us during the project life-time. The final report, which will give an evaluation of this representation on the basis of the experience acquired during the course of the Project, and will present the definitive version of the project computational model, will be delivered at 30 months.

Acknowledgments

Although the report has been prepared mainly at Pisa, all the partners of the Project have contributed substantially: in the first place by representing their own MRDs in the format proposed by Pisa and secondly with much valuable criticism and useful comments. Section 2, in particular, has been prepared taking into account the points raised during a discussion with the other partners at a meeting, in Amsterdam, in March. In addition to the contributions made by the other partners, we must thank the other members of the Pisa group and also Rozemarijn Vogel, student of the Italian Department, University of Amsterdam, actually in Italy under the ERASMUS programme, for all her very useful work in helping to analyze the contents of our machine-readable dictionaries.

SECTION 1

Standardized Representation of the Project Machine-Readable Dictionaries

1.1 Some General Observations on the Representation Model

In this section, we propose a framework which can be used to represent the information present in the different dictionaries that are being analyzed and used in the Project, in a uniform and explicit manner.

The dictionaries which have been examined up to now are *Il Nuovo Dizionario Garzanti*, a monolingual dictionary for Italian, the *Collins Concise English-Italian, Italian-English Dictionary*, the *VanDale Groot Woordenboek Hedendaags Nederlands*, a monolingual Dutch dictionary, the *VanDale Bilingual Dutch/English Dictionary*, the *Longman Dictionary of Contemporary English*, the *Oxford Advanced Learner's Dictionary of Current English*, the *VOX Spanish Monolingual Dictionary*. These dictionaries are here referred to as Garzanti, Collins, VanDale Monolingual, VanDale Bilingual, LDOCE, OALD and VOX, respectively.

At the present stage, the representation is in no way intended to be exhaustive for any particular dictionary type, neither with regard to the types of information that have to be represented, nor to the set of tags which is proposed. Revisions will be made during the course of the Project in order to broaden its coverage*.

The representation consists of three parts: the lexical entry templates, an informal description of the semantics of the Node_tags and the Attribute_tags, and an extensional description of them giving the set of possible values for each Attribute.

1) The Lexical Entry Templates: The Template is a representation of the internal organization of the "maximal" entry in each dictionary, i.e. all possible fields should be foreseen in all possible hierarchically positions. The Template also reflects the hierarchical structure of the dictionary entry; we want to stress the importance of evidencing hierarchical relations among the components of dictionary entries, in contrast with flat tag systems. This hierarchy is best evidenced

---

* Other groups have begun to apply this proposal to their dictionaries and we have already received first positive comments and results in this respect.

by the use of Node_tags, grouping semantically and logically connected constituents (Attribute_tags). Our model is similar to that designed by Neff and Boguraev (1989) for the Longman and Collins dictionaries.

As the Template thus makes transparent the hierarchical structure of the entry and the relations among its components, it will be an essential tool for the parser which has to decode the dictionary entries. In fact, the parser will follow the map of structural relations among the tags in the interpretation of the flat linear dictionary data.

Abbreviated Lexical Entry Templates in which only the main Node_tags appear, without their dependent tags, have also been included for each main template.

2) A list of all the Tags used in all the Templates, and a description of their semantics. This list is not exhaustive, and will be augmented and corrected to include the representation of phenomena or data present in other dictionaries or, if necessary, more refined analyses of the actual dictionaries. The Tags are of two types:
a) Node_Tags, which do not have values, but govern a group of other constituent tags. They have the string 'GROUP' in their name, and are indicated by the name followed by the 'equal' sign and by an informal description.
We have e.g. MORPH_GROUP = node grouping information on....
b) Attribute_Tags, which always take values. They are in lower-case, and are indicated by the name followed by a 'colon' and by a description of the possible values.
We have e.g. morph_label: takes as value any label....
For ease of understanding, at the moment we have used self-explanatory tag-names.

3) The list and description of the possible values which each Attribute_tag can take in each different dictionary, i.e. its domain.
We have not made complete lists for all the Attributes but, for each attribute, we have at least cited some values in order to give an idea of its content.

During the course of the Project, for some of the Attributes (e.g. POS, or certain semantic labels), we may try either to establish a common level for equating the values actually used in the different dictionaries or to standardize the names of the values to be used within the Project.

The following comments are intended as an explanation of particular points of our representation model.

- The optionality and obligatoriness of the Node_ and Attribute_ tags are represented in the Template in the following way:

(...) = optional, and, if present, once only
* = 0 or more times
\+ = 1 or more times
nothing = obligatory, once

If an optional sign is present at the GROUP level, obligatory signs on its constituents are to be intended in the following way: "if the GROUP is present, then this Tag is obligatory".

Given that a generic entry for a dictionary is to be represented, almost all the fields will be optional, because of the wide range of variations in standard dictionary entries.

- The letters 'M' and 'B' at the left of some tags mean that these tags only refer to Monolingual or to Bilingual dictionaries, respectively.

- Hierarchies and dependencies (and therefore scope) are represented by the indentations in the Templates. They may be different in the different dictionaries. This is very important information for parsers. It is our intention to represent these hierarchies formally in the final version of this report.

- Attribute names ending with the string '_type' are used for information which is not explicitly present in written dictionaries, but which can be derived at an early stage with relatively simple automatic or semi-automatic procedures. Additional tags of this kind will be found in the Common Entry, which conforms less to the source text and in which more "derived" information is present (e.g. superordinates or genus-terms).

- Attribute names ending with the string '_text' are used to preserve or point to the data as it is represented in the source text, when the process of parsing the dictionary into

the different fields otherwise results in the source text being no longer recoverable, or when a portion of the source text has not yet been analyzed.
This has been done because maintenance of the source text is considered important: for example, the requirements of further and more refined analyses may necessitate going back to the source.

- We realize that the addition of these '_text' and '_type' attributes may cause the list of tags and the templates to become rather weighty and complicated. Unfortunately, these and other complexities arise from the necessity of disambiguating and representing in an explicit and standardized form, for computational use, texts - in our case dictionaries - which are somewhat inconsistent, incoherent, and not systematic in the way in which they present information, because they are intended for human use. Many difficulties therefore arise if we want to reconcile two, to some extent conflicting, goals: to preserve the original document in its integrity while, at the same time, analyzing its contents in all possible detail. We feel that the efforts made at this stage will in the end pay off by making it possible to have data which is exchangeable and comparable between different research groups and not having to repeat the same time-consuming parsing procedures, in slightly different ways, in different projects.

Finally, it is important to note that what we are presenting in this first section has also to be seen in the perspective of a cooperation with the TEI (Text Encoding Initiative). Therefore, the results of this work are also being used as a contribution to the TEI; we have already had feedback from this group and we expect to have more in the future.

Section 1.2 Lexical Entry Templates

Garzanti: Italian Monolingual
VanDale: Dutch Monolingual
Vox: Spanish Monolingual
OALD: English Monolingual
LDOCE: English Monolingual
Collins: English/Italian Bilingual (Bidirectional)
VanDale: Dutch/English Bilingual (Monodirectional)

TEMPLATE FOR LEXICAL ENTRY

IL NUOVO DIZIONARIO GARZANTI: ITALIAN MONOLINGUAL

Tags at Dictionary Level:

```
DICT_SOURCE
   Dict_name:
   Dict_notes:

   LANGUAGE
      Lang:

   PHONETIC TRANSCRIPTION
        IPA:
```

Tags at Entry Level:

```
   ENTRY
     Entry_Id.:

      HEADWORD_GROUP
          Hdwd_type:
          Hdwd_form:
         (Hdwd_Homonym_no.):
         (VARIANT_GROUP)...

         (PHONETIC_GROUP)
            (Pronunc_text):
            +Pronunciation
         (CROSS-REFERENCE_GROUP)...

         (ETYMOLOGY_GROUP)
             Etymology_text:

         (VARIANT_GROUP)
            (Variant_label):
             Variant_form:
            (PHONETIC_GROUP)...
            (GRAM_INF_GROUP)...

        +HOM_GROUP
             Hom_no.:
            (Hom_form):
            (Hom_compact):
           *(GRAM_INF_GROUP)
              (Gram_Inf_text):

             *(POS_GROUP)
                (POS):
                (subcat):
                (subtype):
                (gender):
                (number):
                (various):

              (INFLECTION_GROUP)
                (Infl_text):
                (Infl_label):
                +Infl_form:
            (VARIANT_GROUP)...
```

```
(CROSS_REFERENCE_GROUP)
      (Xref_text):
      (Xref_type):
       Xref_label:
      +Xref_entry:
      (Xref_extens):

*(SENSE_GROUP)
       Sense_no.:
      (MULTIWORD_GROUP)...
      (CROSS_REFERENCE_GROUP)...

       (DEFINITION_GROUP)
            (GRAM_INF_GROUP)...

           *(SEMANTIC_LABEL_GROUP)
               (Semantic_label_text):
               (Subject_code):
               (Semantic_code):
               (Register_code):
               (Usage_code):
               (Geographic_code):
            (CROSS_REFERENCE_GROUP)...

           *(DEF_GROUP)
               +Def_text:
               (SEMANTIC_LABEL_GROUP)...
              *(CROSS_REFERENCE_GROUP)...
               (TAXONOMY_GROUP)
                  Taxon_label:
                  Taxon_text:

          *(EXAMPLE_GROUP)
             +Ex_text:
             (SEMANTIC_LABEL_GROUP)...
             (Ex_explanation):

          *(MULTIWORD_GROUP)
             (Mwd_label):
              Mwd_form:
             (SEMANTIC_LABEL_GROUP)...
              Mwd_explanation:

               (PROVERB_GROUP)
                  Prov_label:
                 +Prov_text:
                  (Prov_explan):

            (SEMANTIC_RELATIONS_GROUP)

               (SYNONYM_GROUP)
                  Syn_label:
                 +Synonym:
               (ANTONYM_GROUP)
                  Ant_label:
                 +Antonym:
               (ALTERATE_GROUP)
                  Alt_label:
                 +Alterate:

(RUN-ON_GROUP)
   Run-On_type:
   Run-On_label:
   Run-On_form:
```

ABBREVIATED SCHEMA OF THE LEXICAL ENTRY IN GARZANTI

```
ENTRY
   HEADWORD_GROUP
      PHONETIC_GROUP
      ETYMOLOGY_GROUP
      VARIANT_GROUP
      HOM_GROUP
         GRAM_INF_GROUP
            POS_GROUP
            INFLECTION_GROUP
         CROSS_REFERENCE_GROUP
         SENSE_GROUP
             DEFINITION_GROUP
                SEMANTIC_LABEL_GROUP
                DEF_GROUP
                   TAXONOMY_GROUP
                EXAMPLE_GROUP
                MULTIWORD_GROUP
                   PROVERB_GROUP
                SEMANTIC_RELATIONS_GROUP
                   SYNONYM_GROUP
                   ANTONYM_GROUP
                   "ALTERATE"_GROUP
   RUN-ON_GROUP
```

TEMPLATE FOR LEXICAL ENTRY

VANDALE GROOT WOORDENBOEK HEDENDAAGS NEDERLANDS: DUTCH MONOLINGUAL

```
Tags at Dictionary Level:

DICT_SOURCE
   Dict_name:
   Dict_notes:

   LANGUAGE
      Lang:

   PHONETIC TRANSCRIPTION
        IPA:


Tags at Entry Level:

   ENTRY
     Entry_Id:

      HEADWORD_GROUP
          Hdwd_type:
          Hdwd_text:
          Hdwd_form:
         (Hdwd_Homonym_no.):
         (VARIANT_GROUP)

         (HEADWORD_LABEL_GROUP)
            Freq_inf:
          +(SEMANTIC_LABEL_GROUP)...

         (PHONETIC_GROUP)
            (Alternate_pronunc):
            (Stress_position):

         (ETYMOLOGY_GROUP)
             Etymology_text:

         (PROVERB_GROUP)
             Prov_label:

        *(VARIANT_GROUP)
            (Variant_label):
             Variant_form:
            (PHONETIC_GROUP)...
            (GRAM_INF_GROUP)...
            (VARIANT_GROUP)...

         HOM_GROUP
             Hom_no.:

           GRAM_INF_GROUP
              Gram_Inf_text:

               *(POS_GROUP)
                  POS:
                  (subcat):
                  (subtype):
                  (gender):
                  (number):
                  (various):
```

```
    (INFLECTION_GROUP)
      (Infl_text):
      (Infl_label):
      (Infl_stem):
      +Infl_form:
      (CROSS_REFERENCE_GROUP)
         Xref_type:
         Xref_label:
         Xref_text:
         Xref_entry:
         Xref_entry_extens:

*(SENSE_GROUP)
       Sense_no.:

        (DEFINITION_GROUP)
             (GRAM_INF_GROUP)...

            *(SEMANTIC_LABEL_GROUP)
                (Semantic_label_text):
                (Subject_code):
                (Semantic_code):
                (Register_code):
                (Usage_code):
                (Geographic_code):

            *(DEF_GROUP)
                +Def_text:
                (SEMANTIC_LABEL_GROUP)...

             (SEMANTIC_RELATIONS_GROUP)

               (SYNONYM_GROUP)
                  Syn_label:
                 +Synonym:

               (ANTONYM_GROUP)
                  Ant_label:
                 +Antonym:

*(EXAMPLE_GROUP)
    Ex_no.:
   +Ex_text:
   (SEMANTIC_LABEL_GROUP)...
   (Ex_explanation):

  *(COLLOCATION_GROUP)
       Coll_POS:
       Coll_word:

  *(MULTIWORD_GROUP)
       Mwd_label:
       Mwd_form:
       Mwd_explanation:
```

ABBREVIATED SCHEMA OF THE LEXICAL ENTRY IN THE VANDALE MONOLINGUAL

```
DICT_SOURCE
   LANGUAGE
   ENTRY
      HEADWORD_GROUP
         HEADWORD_LABEL_GROUP
         PHONETIC_GROUP
         ETYMOLOGY_GROUP
         PROVERB_GROUP
         VARIANT_GROUP
         HOM_GROUP
            GRAM_INF_GROUP
               POS_GROUP
               INFLECTION_GROUP
            SENSE_GROUP
                DEFINITION_GROUP
                   SEMANTIC_LABEL_GROUP
                   DEF_GROUP
                   SEMANTIC_RELATIONS_GROUP
                      SYNONYM_GROUP
                      ANTONYM_GROUP
            EXAMPLE_GROUP
               COLLOCATION_GROUP
               MULTIWORD_GROUP
```

TEMPLATE FOR LEXICAL ENTRY

VOX: SPANISH MONOLINGUAL

Tags at Dictionary Level:

```
DICT_SOURCE
   Dict_name:
   Dict_notes:

   LANGUAGE
      Lang:
```

Tags at Entry Level:

```
   ENTRY
     Entry_Id.:

      HEADWORD_GROUP
          Hdwd_type:
         (Hdwd_text):
          Hdwd_form:
         (CROSS_REFERENCE_GROUP) ...
         (Hdwd_Homonym_no.):

         (VARIANT_GROUP)
            (Variant_label):
             Variant_form:

         (ETYMOLOGY_GROUP)
             Etymology_text:

             (CROSS_REFERENCE_GROUP)
                   (Xref_text):
                   (Xref_type):
                    Xref_label:
                   +Xref_entry:
                   (Xref_extens):

       +HOM_GROUP
            Hom_no.:
           (Hom_form):
           (Hom_compact):

          *(GRAM_INF_GROUP)
             Gram_Inf_text:

              *(POS_GROUP)
                 (POS):
                 (subcat):
                 (subtype):
                 (gender):
                 (number):
                 (various):

              *(INFLECTION_GROUP)
                 (Infl_text):
                 (Infl_label):
                 +Infl_form:
```

```
+SENSE_GROUP
        Sense_no.:
       (CROSS_REFERENCE_GROUP)...

       (DEFINITION_GROUP)

            *(SEMANTIC_LABEL_GROUP)
                (Semantic_label_text):
                (Subject_code):
                (Semantic_code):
                (Register_code):
                (Usage_code):
                (Geographic_code):
                (Country_code):

            .(DEF_GROUP)
                 Def_text:
                (CROSS_REFERENCE_GROUP)...

           *(EXAMPLE_GROUP)
              +Ex_text:
              (Ex_label):
              (Ex_explanation):

            (MULTIWORD_GROUP)
               Mwd_label:
               Mwd_form:
               DEF_GROUP ...
              (SEMANTIC_LABEL_GROUP)...

            (SEMANTIC_RELATIONS_GROUP)

               (SYNONYM_GROUP)
                  Syn_label:
                 +Synonym:

               (ANTONYM_GROUP)
                  Ant_label:
                 +Antonym:

               ("ALTERATE"_GROUP)
                  Alt_label:
                 +"Alterate":

(SEMANTIC_RELATIONS_GROUP)

(HOMOPH_GROUP)
   Homoph_label:
   Homoph_text:
   Homoph_entry:
```

ABBREVIATED SCHEMA OF THE LEXICAL ENTRY IN VOX

```
ENTRY
   HEADWORD_GROUP
      VARIANT_GROUP
      ETYMOLOGY_GROUP
      HOM_GROUP
         GRAM_INF_GROUP
            POS_GROUP
            INFLECTION_GROUP
         CROSS_REFERENCE_GROUP
         SENSE_GROUP
           CROSS_REFERENCE_GROUP
             DEFINITION_GROUP
                SEMANTIC_LABEL_GROUP
                DEF_GROUP
                EXAMPLE_GROUP
                MULTIWORD_GROUP
                   DEF_GROUP
                   SEMANTIC_LABEL_GROUP
                SEMANTIC_RELATIONS_GROUP
                   SYNONYM_GROUP
                   ANTONYM_GROUP
                   "ALTERATE"_GROUP
   SEMANTIC_RELATIONS_GROUP
   HOMOPH_GROUP
```

TEMPLATE FOR LEXICAL ENTRY

OXFORD ADVANCED LEARNER'S DICTIONARY OF CURRENT ENGLISH: ENGLISH MONOLINGUAL

Tags at Dictionary Level:

```
DICT_SOURCE
   Dict_name:
   Dict_notes:

   LANGUAGE
      Lang:

   PHONETIC TRANSCRIPTION
        IPA:
```

Tags at Entry Level:

```
   ENTRY
     Entry_Id.:

      HEADWORD_GROUP
          Hdwd_type:
          Hdwd_form:
          Hdwd_text:
         (Hdwd_Homonym_no.):

         (PHONETIC_GROUP)
            (Pronunc_text):
            +Pronunciation:

         (VARIANT_GROUP)
            (Variant_label):
             Variant_form:

        +HOM_GROUP
             Hom_no.:
            (Hom_form):

           *(GRAM_INF_GROUP)
              (Gram_Inf_text):

               *(POS_GROUP)
                  (POS):
                  (pos_subcat):
                  (pos_subtype):
                  (pos_gender):
                  (pos_number):
                  (pos_various):
                  (pos_gcode):

                (INFLECTION_GROUP)
                  (Infl_text):
                  (Infl_label):
                  +Infl_form:

                AUX_ GROUP...
                  Aux_label:
                  Aux_form:
```

```
*(SENSE_GROUP)
        Sense_no.:

        (DEFINITION_GROUP)
             (GRAM_INF_GROUP)...

              SENSE_LABEL_GROUP

            *(SEMANTIC_LABEL_GROUP)
                (Semantic_label_text):
                (Subject_code):
                (Semantic_code):
                (Register_code):
                (Usage_code):
                (Geographic_code):

            *(DEF_GROUP)
               +Def_text:
               (SEMANTIC_LABEL_GROUP)...
               (CROSS_REFERENCE_GROUP)

            *(EXAMPLE_GROUP)
                +Ex_text:
                (Ex_explanation):

            *(MULTIWORD_GROUP)
                 Mwd_label:
                 Mwd_form:
                 Mwd_explanation:

            (SEMANTIC_RELATIONS_GROUP)

               (ANTONYM_GROUP)
                  Ant_label:
                 +Antonym:

CROSS_REFERENCE_GROUP
  Xref_label:
  Xref_text:
  Xref_type:
  Xref_entry:
  Xref_entry extens.:

(RUN-ON_GROUP)
  Run-On_type:
  Run-On_label:
  Run-On_form:
```

ABBREVIATED SCHEMA OF THE LEXICAL ENTRY IN OALD

```
ENTRY
   HEADWORD_GROUP
      PHONETIC_GROUP
      VARIANT_GROUP
      HOM_GROUP
         GRAM_INF_GROUP
            POS_GROUP
            INFLECTION_GROUP
            AUX_GROUP
         SENSE_GROUP
             DEFINITION_GROUP
                SENSE_LABEL_GROUP
                SEMANTIC_LABEL_GROUP
                DEF_GROUP
                EXAMPLE_GROUP
                MULTIWORD_GROUP
                SEMANTIC_RELATIONS_GROUP
                   ANTONYM_GROUP
   CROSS_REFERENCE_GROUP
   RUN-ON_GROUP
```

TEMPLATE FOR LEXICAL ENTRY

LONGMAN DICTIONARY OF CONTEMPORARY ENGLISH: ENGLISH MONOLINGUAL

Tags at Dictionary Level:

```
DICT_SOURCE
  Dict_name:
  Dict_notes:

  LANGUAGE
    Lang:

  PHONETIC_TRANSCRIPTION
    IPA:
```

Tags at Entry Level:

```
ENTRY
  Entry_Id:

  HEADWORD_GROUP
    Hdwd_type:
    Hdwd_form:
    Hdwd_text:
   (Hdwd_homonym_no):
   (VARIANT_GROUP)...

  (PHONETIC_GROUP)
    (Pronunciation_text):
     +Pronunciation:

  VARIANT_GROUP
   (Variant_label):
    Variant_form:
    (PHONETIC_GROUP)...
    (GRAM_INF_GROUP)...

  +HOM_GROUP
    Hom_no:
    (Hom_form):
    (Hom_compact):

  *(GRAM_INF_GROUP)
      (Gram_inf_text):
      +POS_GROUP
        POS:
        (Pos_subtype):
       *(gcode):

    *(INFLECTION_GROUP)
        Infl_text:
       (Infl_label):
        Infl_form:


    *(AUX_GROUP)
        Aux_label:
        Aux_form:
```

```
*(SENSE_GROUP)
  Sense_no:
  DEFINITION_GROUP
    (GRAM_INF_GROUP)...

     (SEMANTIC_LABEL_GROUP)
       (Semantic_label_text):
       (Subject_code):
       (Semantic_code):
       (Register_code):
       (Usage_code):
       (Geographic_code):
       (VARIANT_GROUP)...

       DEF_GROUP
        Def_text:

          (PROVERB_GROUP)
          (Prov_label):
           Prov_text:
           Prov_explanation:
          (PHONETIC_GROUP)...

          (CROSS_REFERENCE_GROUP)
            (Xref_text):
            (Xref_type):
             Xref_label:
            +Xref_entry:
           *(Xref_extens):
            (SENSE_LABEL_GROUP)...

         (TAXONOMY_GROUP)
           (Taxon_label):

        *(EXAMPLE_GROUP)
           Ex_text:
           Ex_explanation

      (SEMANTIC_RELATIONS_GROUP)

         (SYNONYM_GROUP)
           Syn_label:
           Syn-hdwd:

         (ANTONYM_GROUP)
           Ant_label:
           Ant_hdwd:

         (SEMANTIC_ROLES_GROUP)
           Arg1:
          (Arg2):
          (Arg3):

      (MULTI_WORD_GROUP)
         Mwd_label:
         Mwd_form:
         Mwd_explanation:

(RUN_ON_GROUP)
     (Run_on_label):
      Run_on_form:
     (GRAM_INF_GROUP)...
```

ABBREVIATED SCHEMA OF THE LEXICAL ENTRY IN LDOCE

ENTRY

  HEADWORD_GROUP

    PHONETIC_GROUP

    VARIANT_GROUP

    HOM_GROUP

      GRAM_INF_GROUP

        POS_GROUP

        INFLECTION_GROUP

        AUX_GROUP

      SENSE_GROUP

        DEFINITION_GROUP

          SEMANTIC_LABEL_GROUP

            DEF_GROUP

              PROVERB_GROUP

              CROSS_REFERENCE_GROUP

              TAXONOMY_GROUP

                EXAMPLE_GROUP

                SEMANTIC_RELATIONS_GROUP

                  SYNONYM_GROUP

                  ANTONYM_GROUP

                  SEMANTIC_ROLES_GROUP

          MULTI_WORD_GROUP

    RUN_ON_GROUP

TEMPLATE FOR LEXICAL ENTRY
COLLINS BILINGUAL: ENGLISH/ITALIAN, ITALIAN/ENGLISH

Tags at Dictionary Level:

```
DICT_SOURCE
   Dict_name:
   Dict_notes:

   LANGUAGE
        L1:
        L2:
        Metalanguage:

   PHONETIC TRANSCRIPTION
        IPA:
        Notes:
```

Tags at Entry Level:

```
   ENTRY
     Entry_Id.:

      HEADWORD_GROUP
          Hdwd_type:
         (Hdwd_text):
          Hdwd_form:
         (Hdwd_Homonym_no.):

            (PHONETIC_GROUP)
               (Pronunc_text):
               +Pronunciation:

            (CROSS_REFERENCE_GROUP)...

            (VARIANT_GROUP)
               (Variant_label):
                Variant_form:
               (PHONETIC_GROUP)...
               (GRAM_INF_GROUP)...

           *(INFLECTION_GROUP)...
                (PHONETIC_GROUP)...

            (CROSS_REFERENCE_GROUP)...

           +HOM_GROUP
               Hom_no.:
               (Hom_form):
                  (PHONETIC_GROUP)...
               (Hom_compact):

            (CROSS_REFERENCE_GROUP)...

             (MULTIWORD_GROUP)
                Mwd_label:
               +Mwd_form:

               (GRAM_INF_GROUP)
                    (Gram_Inf_text):
```

```
        *(POS_GROUP)
             POS:
             (subcat):
             (subtype):
             (gender)
             (number):
             (various):
         (INFLECTION_GROUP)
            (Infl_text):
           *(Infl_label):
            +Infl_form:
            (PHONETIC_GROUP)...
         (AUX_GROUP)
             Aux_label:
             Aux_text:
    (CROSS_REFERENCE_GROUP)...

    *(SENSE_GROUP)
        Sense_no.:
        (Hom_form):
        (GRAM_INF_GROUP)...
        (COMPOUND_GROUP)...

        (TRANSLATION_GROUP)
           .(SENSE_LABEL_GROUP)
                (Sense_Label_text):
                (GRAM_INF_GROUP)..
               *(SEMANTIC_LABEL_GROUP)
                  (Subject_code):
                  (Semantic_code):
                  (Register_code):
                  (Usage_code):
                  (Geographic_code):

               *(SEMANTIC_INDICATOR_GROUP)
                  (Semantic_Indicator_type):
                   Semantic_Indicator_text:

            *(TRANS_GROUP)
               (Trans_type):
               (Trans_label):
               +Trans_text:
               (GRAM_INF_GROUP)...

            *(EXAMPLE_GROUP)
               +Ex_text:
              *(SENSE_LABEL_GROUP)...

               +EXAMPLE_TRANS_GROUP
                    (Ex_Trans_type):
                    (Ex_Trans_label):
                    +Ex_Trans_text:
                       (GRAM_INF_GROUP)...

    (CROSS_REFERENCE_GROUP)
       (Xref_text):
       (Xref_type):
        Xref_label:
        Xref_entry:
       (Xref_entry_extens):

(RUN-ON_GROUP)
   Run-On_type:
   Run-On_label:
   Run-On_form:
```

ABBREVIATED SCHEMA OF THE LEXICAL ENTRY IN COLLINS

```
ENTRY
   HEADWORD_GROUP
          PHONETIC_GROUP
          VARIANT_GROUP
          HOM_GROUP
          MULTIWORD_GROUP
             GRAM_INF_GROUP
                 POS_GROUP
                 INFLECTION_GROUP
                 AUX_GROUP
              SENSE_GROUP
                 TRANSLATION_GROUP
                     SENSE_LABEL_GROUP
                         SEMANTIC_LABEL_GROUP
                         SEMANTIC_INDICATOR_GROUP
                      TRANS_GROUP
                      EXAMPLE_GROUP
                        EXAMPLE_TRANS_GROUP
            CROSS_REFERENCE_GROUP
   RUN-ON_GROUP
```

TEMPLATE FOR LEXICAL ENTRY
VANDALE BILINGUAL: DUTCH/ENGLISH

Tags at Dictionary Level:

```
DICT_SOURCE
   Dict_name:
   Dict_notes:

   LANGUAGE
        L1: Dutch
        L2: English
        Metalanguage: Dutch (monodirectional)
```

Tags at Entry Level:

```
   ENTRY
     Entry_Id.:

      HEADWORD_GROUP
          Hdwd_type:
          Hdwd_text:
          Hdwd_form:
         (Hdwd_Homonym_no.):

         (HEADWORD_LABEL_GROUP)
          *(SEMANTIC_LABEL_GROUP)...

         (PROVERB_GROUP)
           Prov_label:

        HOM_GROUP
             Hom_no.:

           GRAM_INF_GROUP
              Gram_Inf_text:

               *(POS_GROUP)
                  POS:
                  (subcat):
                  (subtype):
                  (gender):
                  (number):
                  (various):

            *(SENSE_GROUP)
                    Sense_no.:

                    (GRAM_INF_GROUP)...

                    (TRANSLATION_GROUP)

                        (SENSE_LABEL_GROUP)
                            (Sense_Label_text):

                           *(SEMANTIC_LABEL_GROUP)
                              (Subject_code):
                              (Semantic_code):
                              (Register_code):
                              (Usage_code):
                              (Geographic_code):
```

```
                    *(SEMANTIC_INDICATOR_GROUP)
                      (Semantic_Indicator_type):
                       Semantic_Indicator_text:

                *(TRANS_GROUP)
                    (Trans_type):
                    (Trans_label):
                    +Trans_text:
                    (GRAM_INF_GROUP)...
                   *(SEMANTIC_LABEL_GROUP)...
                   *(SEMANTIC_INDICATOR_GROUP)...
                   *(TRANS_GROUP)...

*(EXAMPLE_GROUP)
    Ex_no.:
   +Ex_text:
  *(SEMANTIC_LABEL_GROUP)...
  *(SEMANTIC_INDICATOR_GROUP)...

  *(COLLOCATION_GROUP)
      Coll_type:
      Coll_text:
  *(MULTIWORD_GROUP)
      Mwd_label:
      Mwd_form:
      Mwd_explanation:

   +EXAMPLE_TRANS_GROUP
      (Ex_Trans_type):
      (Ex_Trans_label):
      +Ex_Trans_text:
     *(SEMANTIC_LABEL_GROUP)...
     *(SEMANTIC_INDICATOR_GROUP)...
```

ABBREVIATED SCHEMA OF THE LEXICAL ENTRY

VANDALE BILINGUAL DUTCH-ENGLISH

```
DICT_SOURCE
   LANGUAGE
   ENTRY
      HEADWORD_GROUP
         HEADWORD_LABEL_GROUP
         HOM_GROUP
            GRAM_INF_GROUP
               POS_GROUP
            SENSE_GROUP
                  TRANSLATION_GROUP
                     SENSE_LABEL_GROUP
                        SEMANTIC_LABEL_GROUP
                        SEMANTIC_INDICATOR_GROUP
                     TRANS_GROUP
                        SEMANTIC_LABEL_GROUP
                        SEMANTIC_INDICATOR_GROUP
                        TRANS_GROUP
            EXAMPLE_GROUP
               SEMANTIC_LABEL_GROUP
               SEMANTIC_INDICATOR_GROUP
               COLLOCATION_GROUP
               MULTIWORD_GROUP
               EXAMPLE_TRANS_GROUP
                   SEMANTIC_LABEL_GROUP
                   SEMANTIC_INDICATOR_GROUP
```

1.3 Explanation of Lexical Entry Template Tags and Description of Attribute Values

Tags at Dictionary Level are:

DICT_SOURCE = node grouping information on the source of the lexical data.
Dict_Name: the name of the source dictionary.
Dict_Notes: contains information on the kind of lexical data contained in the source dictionary. For example the VanDale material originates from a historically oriented database, but is restricted to Contemporary Dutch. It includes words from spoken language but excludes dialect words. Entries are based on a "one-form-one-entry" principle.

LANGUAGE = node grouping information on language of source dictionary.

M **Lang**: language of source dictionary for monolingual dictionaries.
B **L1** : source language for bilingual dictionaries.
B **L2** : target language for bilingual dictionaries.
e.g. for an Italian/English bilingual dictionary: in the Italian-English dataset, L1=Italian, L2=English; in the English-Italian dataset, L1=English, L2=Italian.

B **Metalanguage**: Language used as metalanguage. In bidirectional bilingual dictionaries, this is normally L1 for each dataset whereas, in monodirectional bilinguals, the metalanguage used is normally that of the intended user.
For example, in our Collins bilingual, in the Italian-English dataset, this field takes as value Italian and, in the English-Italian dataset, the value is English; in the VanDale bilingual which is monodirectional, the language used as metalanguage is always Dutch.

PHONETIC TRANSCRIPTION

IPA: Symbols of the International Phonetic Association; takes as values Yes/No.

Notes: comments on particular characteristics of phonetic transcription used by source dictionary.
For example, in Collins, OALD and LDOCE the phonetic transcription includes information on word stress.

Tags at Entry Level are:

ENTRY = node for the entire lexical entry.

Entry_Id: takes as value a number identifying the entry on the source tape in order to preserve the source order (e.g. when entries are added, deleted, split, etc., when loading dictionary into the LDB shell, or when transferring information to other projects which have also used the same source tape).

HEADWORD_GROUP = node grouping information on headword.

Hdwd_type: all "_type" tags contain information which is implicit in the entry and can be derived either manually or automatically. This tag indicates the particular kind of headword and can take as values, e.g. lemma, irregular word-form, suffix, prefix, abbreviation, etc.
A list of possible labels is given in Section 1.4.

Hdwd_text: This field is necessary to preserve the source text in at least two cases:
a) headwords in which hyphenation information is included are transcribed here, exactly as they appear. For example, in the Collins Bilingual, the English-Italian dataset includes hyphenation information in the headword: "in.for.ma.tion".
b) headwords in which morphological information is located in the headword position are transcribed here, exactly as they appear.
For example, in the Collins Bilingual, the Italian-English dataset includes many headwords like "basico,a,ci,che": the whole string is entered in this field, the form assumed as the primary form ("basico" in the example here) will be entered as value in the Hdwd_form field and "a,ci,che" will be entered as value in the Infl_text field of the Inflection_Group, which in this case will substitute the Variant_Group.

Hdwd_form: Obligatory field; takes as value the dictionary citation form. Contains only the primary form for "complex" headwords, see above Hdwd_text. All signs which are additional to the actual graphic form of the headword must be removed, e.g.indications of hyphenation, stress, etc., whereas all signs which belong to the usual graphic form of the headword must be maintained, e.g. capitals, graphic accents, periods, spaces, etc.(for example, Cristo, wagon-lit, ab aeterno).

Hdwd_Homonym_no.: Usually takes as value a number (sometimes this appears in dictionaries as a superscript number) or any other sign used by the source dictionary to represent homonyms recorded as separate entries. Collins and Garzanti, at times, indicate both lexical and grammatical homonyms in this way, e.g. for "calcio" both give separate entries for 2 lexical homonyms (both nouns), whereas for "potere" separate entries are given for grammatical homographs (noun and verb).

HEADWORD_LABEL_GROUP = node grouping information related to the entry as a whole and not to specific senses. In VanDale all values from the Semantic_Label_Group can occur at this level.

Freq_Inf: tag containing frequency information. The Van Dale dictionary tape, for example, has frequency labels which either refer to all senses or for which it is not known to which sense they apply.

PHONETIC_GROUP = node grouping phonetic information.This node can be repeated in different positions in the lexical entry.

Pronunc_text: Takes as value the entire pronunciation field as it appears in the dictionary, without any analysis. This field is used to preserve the source text, e.g. when more than one pronunciation is merged in a single string. For example, in Collins, for the entry "fuor(i)uscito", the two pronunciations are indicated in the same way as the headword , i.e. with the "i" between brackets, and will thus be recorded in two different Phonetic_groups, one for the headword, one for the variant.

Pronunciation: takes as value the pronunciation information. When more than one pronunciation is associated with the headword or with any variant form, this field is repeated; the first occurrence is assumed as the primary pronunciation.
For example, Garzanti gives "nesso [nès o nés]"; in this case, the entire pronunciation is entered in the Pronunc_text field, "nès" is entered in a first Pronunciation field and "nés" is entered in a second field.

Primary_stress_pos: takes as value a number indicating character position in the headword of the primary stress.

Secondary_stress_pos: takes as value a number indicating character position in the headword of the secondary stress.

VARIANT_GROUP = node grouping information on variant forms of the headword.
When inflected word-forms are recorded at headword level, (see the example "basico, a, ci, che" in the Hdwd_text field above), the Inflection_group node substitutes the Variant_Group node.

Variant_label: takes as value any label given in the source dictionary indicating the type of variant.
The list of possible labels for our dictionaries is given in Section 1.4.

Variant_form: takes as value the graphical form of the variant as it appears in the dictionary,
e.g. Collins gives "sceptic, (Am) skeptic [.....]";
"skeptic" will be entered in the Variant_form field;
in the example of the entry "fuor(i)uscito" cited above,
"fuoruscito" is entered as value in the Hdwd_form field
and "fuoriuscito" is entered as value in this field.

HOM_GROUP = node for each homograph group within an entry (usually grammatical homographs). Homograph divisions are usually given for major POSs, but separate homographs are also often given for different subcategorizations of the same POS, e.g. vt, vi, vr, etc..
When there is just one homograph group, it is not necessarily indicated explicitly in the dictionary, but is logically always present.

Hom_no.: takes as values the number or label given in the dictionary, or the value NIL when there is no explicit number or label.
The list of possible labels for our dictionaries is given in Section 1.4.

Hom_form: this field is necessary whenever the form of the entry word at the homographic level differs from the form which appears at the headword level.
For example, in the Italian-English dataset, Collins lists the reflexive form of verbs in a separate homograph division and indicates their graphic form; e.g. in the entry for "abbandonare", there are two Homograph groups for the transitive and the reflexive forms of the verb, as follows:
1 vt ... 2: ~rsi vr.
In this case, Hom_Form takes the value "~rsi".

Hom_compact: takes as value Yes; this field is present only when two or more primary POSs are listed under one Hom_No.
It will be disambiguated to create two or more entries or homographs in the Common Lexical Entry.
For example, Garzanti gives "ambulante agg. e s.m. e f.";
in Collins, we have "Coreano, a [....] ag, sm, sm/f".

GRAM_INF_GROUP = node grouping morphological and syntactic information. This node can be repeated in many different positions in the lexical entry.

Gram_Inf_text: field needed to preserve source text. For example, in Collins, for the headword "behove" we find "impers vt"; thus, in order to maintain the order, "impers vt" will be entered in the "Gram_Inf_text" field, "v" in the "POS" field, "t" in the "subcat" field and "impers" in the "subtype" field.
It is also used for complex grammatical information which will need further analysis in the future.

POS_GROUP = node grouping all information on parts-of-speech.

POS: takes as values the major parts-of-speech (verb, noun, etc.) as they are represented in each dictionary.
The list of possible labels for our dictionaries is given in Section 1.4.
In the future, either a standard list or conversion tables for the individual lists of each source dictionary will be defined for the project.

subcat: takes as value any subcategorization information given for verbs, e.g.on transitivity,reflexivity, etc.
The list of possible labels (plus examples) for our dictionaries is given in Section 1.4.

subtype: takes as value any labels which regard more specific grammatical information for any POS.
The list of possible labels (plus examples) for our dictionaries is given in Section 1.4.

gender: takes as value any labels for nouns, pronouns and adjectives regarding gender.
The list of possible labels (plus examples) for our dictionaries is given in Section 1.4.

number: takes as value any labels for nouns, pronouns and adjectives regarding number.
The list of possible labels (plus examples) for our dictionaries is given in Section 1.4

various: this field is used at the moment for "junk collection", i.e. all grammatical information which cannot be classified under any of the previous headings. For example, in Collins in the entry for "be", aux verb, we find "with prp: forming continuous tenses"; at this stage we put this information in the various field.

g-code: this field is used for codes such as the Longman formal grammatical codes which describe in detail the syntactic behaviour of headwords or their word senses.

INFLECTION_GROUP = node grouping information on inflection, usually gives irregular inflected forms.
As has been explained in the hdwd_text description, it has also been found necessary at times to use this node for morphological information located in the headword position; in this case, this node will be located in the position of the Variant_Group.

Infl_text: This field is necessary to preserve the source text and takes as value the morphological information, as it is reported in the dictionary, when it cannot be analyzed.
For example, in Garzanti, we find "ciondolare ... v.intr.(io ciondolo ecc.)"; in this case "io ciondolo ecc." will be entered as value in this field.

Infl_label: takes as value any labels indicating the type of morphological information given.
The list of possible labels for our dictionaries is given in Section 1.4.

Infl_stem: takes as value the "stem" of the entry lemma. For example, Dutch needs a specification of the stem-form for flection and derivation.

Infl_form: takes as value inflected word forms or inflectional endings, as they appear in the dictionary. For example, in Garzanti, we find "doppiatore...s.m. (f.-trice)"; in this case, "f" will be entered as value in the Infl_label field and "-trice" in this field.

AUX_GROUP = node grouping information on the auxiliary verb associated with a given homograph of the headword, according to its subcategorization.

Aux_label: takes as value any label indicating the presence of an auxiliary verb.
The list of possible labels for our dictionaries is given in Section 1.4.

Aux_form: takes as value the auxiliary verb given.

SENSE_GROUP = node grouping the information on each sense division of the headword. Usually contains labels, definitions and examples for monolingual dictionaries, and labels, translations and their examples for bilinguals.

Sense_no.: takes as value the number, letter, etc., which is used in the entry to distinguish between the different word senses. For monosemous words, the dictionary sense_no. may be implicit. The value in this case is NIL.
The list of possible labels for our dictionaries is given in Section 1.4.

M **DEFINITION_GROUP** = node grouping information on the meaning of the headword, normally contains semantic labels and definition(s).(For bilingual dictionaries, in this position we have the Translation_group.)

SENSE_LABEL_GROUP = node grouping explicit and implicit information on the sense being defined or translated. (This node governs the two following nodes and is repeated in all the different positions in which such information appears.)

Sense_label_text: field needed to preserve source text for the sense labels, both with regard to the order of the labels and to handle information which can be found in this field in addition to the codes and will require further analysis.
For example, in Collins, under the first sense division of the English-Italian entry for "superior", we find the sense labels (Comm: goods, quality) which will eventually have to be disambiguated.

SEMANTIC_LABEL_GROUP = node grouping one or more explicit semantic labels attached to a word sense. For each of the Semantic_label types below, the list of possible labels for our dictionaries is given in Section 1.4. Where possible, in the future,either a standard list or conversion tables for the particular lists of each source dictionary will be defined for the project.

Semantic_label_text: field needed to preserve source text for semantic labels, both with regard to the order of the labels and to handle information which can be found in this field in addition to the codes and will require further analysis.
For example, in Garzanti, after the definition of "barricarsi" we find "(anche fig.)" where the word "anche = also" is not a code, but is information which we do not wish to lose.

Subject_code: labels indicating particular domains. Values for each of our dictionaries are listed in Section 1.4.

Semantic_code: labels indicating metaphorical, figurative usage, etc.
Values for each of our dictionaries are listed in Section 1.4.

Register_code: labels indicationg colloquial, formal, informal, literary, poetical usage, etc.
Values for each of our dictionaries are listed in Section 1.4.

Usage_code: labels indicating archaic, old, rare usage, etc.
Values for each of our dictionaries are listed in Section 1.4.

Geographic_code: labels indicating dialectal, regional usage, etc.
Values for each of our dictionaries are listed in Section 1.4.

Country_code: labels indicating national variants. This label is used, for example, by Vox for American variants of Spanish. See Section 1.4.

B **SEMANTIC_INDICATOR_GROUP** = node grouping information on semantic indicators or constraints on the translations.

Semantic_Indicator_type: identifies the particular kind of semantic indicator used, and can take values such as "near-synonym", "contextual", etc. The values for our dictionaries are listed in Section 1.4.

Semantic_Indicator_text: takes as value the semantic indicator information as it appears in the source dictionary.

M **DEF_GROUP** = node grouping definitions regarding word senses.

Def_text: takes as value the definitions as they appear in the source dictionary without any analysis.

M **TAXONOMY_GROUP** = node grouping information on typical taxonomies (usually for animals and plants),if explicitly given in the dictionary.

Taxon_label: takes as value the label indicating the taxonomy type.
The list of possible indicators for this attribute in our dictionaries is given in Section 1.4.

Taxon_text: takes as value what is given in the source dictionary as taxonomy data.

B **TRANSLATION_GROUP** = node grouping all the information for each word sense, including translations, sense_labels, examples, etc. this node corresponds to the Definition_ Group for monolingual dictionaries.

B **TRANS_GROUP** = node grouping the translations for each word-sense.

Trans_type: indicates the particular kind of translation given, when it is not a direct L2 equivalent but e.g. an explanation, a cultural equivalent, etc. The values for our dictionaries are listed in Section 1.4.

Trans_label: any label indicating explicitly the trans_type.
The values for our dictionaries are listed in Section 1.4.

Trans_text: takes as value the translation(s) as they appear in the source dictionary.
Gram_Inf regarding the trans_text may appear at the end of or within the Trans_text; in the latter case the Trans_text will continue in a **Trans_text_cont** field.

EXAMPLE_GROUP = node grouping examples referring to a word sense.

Ex_no.: takes as value a number identifying the example. This is needed by the VanDale dictionaries, in which the examples are stored separately from the senses, in order to link examples to the related sense.

Ex_label: takes as value any label indicating the presence of any example.
The list of values for our dictionaries is given in Section 1.4.

Ex_text: takes as value the example(s) as they appear in the source dictionary.

Ex_explanation: takes as value any explanation, gloss or paraphrase which is associated with a particular example.
For example, in Garzanti, under the headword "legno" (wood), we find the example "legno compensato" (plywood) followed by an explanation "formato di piu' strati di diversi qualita'" (formed of several layers of different qualities)
In Longman, under the headword "foreign", there is the example "The swelling on her finger was caused by a foreign body in it" followed by the explanation "(= a small piece of some solid material that had entered it by accident)".

B **EXAMPLE_TRANS_GROUP** = node grouping information on the translation(s) of an example.

Ex_trans_type: indicates the particular kind of translation given for an example when it is not a direct translation of the example, but e.g. an explanation, a cultural equivalent, etc.
The values for our dictionaries are listed in Section 1.4.

Ex_trans_label: any label indicating explicitly the Ex_trans_type.
The values for our dictionaries are listed in Section 1.4.

Ex_trans_text: takes as value the translation(s) of an example as it appears in the source dictionary.
Gram_Inf regarding the Ex_Trans_text may appear at the end of or within the Ex_Trans_text; in the latter case the Ex_trans_text will continue in an **Ex_trans_text_cont** field.

COLLOCATION_GROUP = node grouping information on typical collocations of the headword in a particular sense with words with particular parts of speech.

Coll_POS: this field contains information on the POS of the word with which the headword collocates.

Coll_word: this field contains the actual collocation word.

PROVERB_GROUP = node grouping proverbs referring to a headword or to a word sense.

Prov_label: takes as value the label indicating the presence of a proverb.
The list of possible indicators for this attribute in our dictionaries is given in Section 1.4.

Prov_text: takes as value any proverb(s).

Prov_explanation: takes as value any explanations which are associated with a particular proverb.

SEMANTIC_RELATIONS_GROUP = node grouping information on particular explicit semantic relations.

SYNONYM_GROUP = groups any explicit information concerning synonyms of the headword or of a word sense..

Syn_label: takes as value the label indicating the presence of synonym(s).
The values for each of our dictionaries are listed in Section 1.4.

Synonym: takes as value the synonym(s).

ANTONYM_GROUP = groups any explicit information concerning antonyms of the headword or of a word sense.

Ant_label: takes as value the label indicating the presence of antonyms.
The values for each of our dictionaries are listed in Section 1.4.

Antonym: takes as value the antonym(s).

ALTERATE_GROUP = groups any explicit information concerning words such as diminutives, augmentatives, pejoratives, etc.

Alt_label: takes as value the label indicating the presence of "alterates".
The values for each of our dictionaries are listed in Section 1.4.

Alterate: takes as value explicit "altered forms" of the headword.

SEMANTIC_ROLES_GROUP = node grouping explicit information concerning the predicate argument structure of the verb. For example, see the box codes which are present on the Longman dictionary tape although they are not given in the printed volume.

Arg1: the value is taken from the corresponding LDOCE box code.

Arg2: see above

Arg3: see above

MULTIWORD_GROUP = node grouping explicit information on various types of multiple words, e.g. phrasal verbs, frozen expressions, etc., listed within the entry.
This node can appear in different positions, depending on the dictionary. For example, in Collins, this node is used for compounds and is given the same status as a homograph division, however, it can contain more than one compound form, each of which can have its own Gram_Inf, Sense_Labels and Trans_Groups. In Garzanti, this node contains compounds, idioms, fixed phrases, etc., and is located either immediately after the Sense_no. or within the Sense_Group, after the Example_Group and the whole node is repeated with its Semantic_Labels for each multiword set.

Mwd_label: takes as value the label indicating the presence of some kind of multiple word.
The list of possible labels (plus examples) for our dictionaries is given in Section 1.4.

Mwd_form: takes as value the multiword form(s) given in the dictionary.

Mwd_explanation: takes as value any explanation which is associated with a particular Mwd_form.

ETYMOLOGY_GROUP = node grouping information on etymology.

Etymology_text: takes as value the etymology information reported in the dictionary without any analysis.

CROSS-REFERENCE_GROUP = node grouping information on cross-references of various types. This node can appear at almost any level in the lexical entry.

Xref_type:cross-references can be of various types, both explicit and implicit, to other dictionary entries or to illustrations.The values which can be entered in this field are thus "explicit", "implicit", "fig".
For example, Vox gives references to figures at the headword level. In this case, "fig" will be entered in this field. Often, definitions contain implicit cross-references: for example, in Garzanti, we find "embolia.. s.f. ostruzione di ... causata da ... un corpo estraneo (embolo)"; in this case, "implicit" is entered as value in this field and "embolo" is entered in the Xref_entry field.
The list of possible labels for our dictionaries is given in Section 1.4.

Xref_label: takes as value the label, sign, or word(s) used in the dictionaries to indicate that a cross-reference follows.
The list of possible labels for our dictionaries is given in Section 1.4.

Xref_text: takes as value the entire cross-reference information in order to preserve the source text, when necessary.
For example, Garzanti gives "lagrima... -> lacrima e deriv.", where the words "e deriv." (= and its derivatives) need further analysis.

Xref_entry: takes as value the entry which is cross-referenced.

Xref_entry_extens: takes as value any Homonym, Homograph and/or Sense nos. qualifying the referenced entry.
For example, in Collins we find "bale [...] vt,vi see bale out 1,2(a)" where "see" is entered in the Xref_label field, "bale out" is entered in the Xref_entry field, and "1,2(a)" is entered in the Xref_entry_extens field.

RUN-ON_GROUP = node grouping run-ons to the headword; the main feature of such run-ons is formal, i.e. they are strictly related to the headword, e.g. derivatives formed from the headword by adding a suffix, phrasal verbs, compounds, etc., and have not been given independent entry status in the particular dictionary under analysis but have been located at the end of the entry as a type of sub-entry. The structure of this node is often similar to the structure of the entire entry node.
For example, in Garzanti, the structure of this node is identi cal to that of the Headword_Group, beginning with the Hom_Group, etc. after the Run-On_form.
For Collins, the structure is the same as for Garzanti.
In Longman, the information given here is on derivatives and the node ends with the Gram_Inf_Group.
In Oxford, this group contains compounds, derivatives, phrasal verbs and idioms; the structure is similar to that of the entry.

Run-On_type: identifies the particular kind of run-on and can take values such as "suffix", "phrasal verb",etc..
The values for each of our dictionaries are listed in Section 1.4.

Run-On_label: takes as value the label, sign, or word(s) used in the dictionaries to indicate that a run-on follows.
The list of possible labels for our dictionaries is given in Section 1.4.

Run-On_form: takes as value the derived word, the suffix used to form the derived word, the phrasal verb, etc. as they appear in the dictionary.
For example, Garzanti gives "esterno ... agg. ... //-mente avv. all'esterno ..."; "-mente" will be entered as value in this field.

HOMOPH_GROUP = node grouping homophones of the headword (i.e. forms with the same pronunciation as the headword but with a different orthographic form).
For example, in Vox the entry "cabe" has "cave" as homophone.

Homoph_label: takes as value the label used in the dictionary to indicate the presence of homophone(s). The list of possible labels for our dictionaries is given in Section 1.4.

Homoph_text: takes as value the whole text corresponding to this field.

Homoph_entry: takes as value the list of entries which are referenced.

1.4 List of Values for Template Attributes in the Project Machine-Readable Dictionaries

The values for the various attributes in the Lexical Entry Template which are found in the machine-readable dictionaries used so far within the Project are listed here.
We have not givem complete lists for all the Attributes but we have listed at least some values for each Attribute in order to give an idea of its content.It will only be possible to have completely exhaustive lists for each dictionary when the first parsing and analysis stages have been completed.

Hdwd_type:

Possible values: lemma (typical standard dictionary headword)
word-form (inflected word-forms, usually irregular)
abbreviation
acronym
suffix
prefix
proper noun
contracted form (e.g. ain't)
compound (e.g. bathing costume)
phrasal verb
letter

in Vox
appocopation
onomatopoeia

Hdwd_Homonym_No.:

in Garzanti:
NIL (if there is only one homonym)
superscript 1,2,3,...n (otherwise)

in Collins:
NIL (if there is only one homonym)
superscript 1,2,3,...n (otherwise)

in VanDale:
NIL (if there is only one homonym)
1,2,3,...n (otherwise)

in Vox:
NIL (if there is only one homonym)
1,2,3,...n (otherwise)

Variant_label:

in Garzanti: raro = rare
.....

in Collins: Am = American
Brit = British
Scot = Scottish
.....

in VanDale: not present.

in LDOCE: note on restriction of distribution (e.g. a-an, before a vowel sound)

in Vox: VAR, var, Tambien = variant
Incorrecto, es incor. = incorrect variant
"-" followed by a solid suffix, in the Hdwd_text could indicate a variant, although, usually, it is a morphological group.

Hom_No.:

in Garzanti: NIL if there is only one homograph
// otherwise; however, when there is more than one homograph, Garzanti only indicates them specifically with this label from the second on,
e.g. **dopopranzo** *s.m.....// avv.....*
appiccicare *v.tr...//v.intr..//-arsi v.rifl....*

in Collins: NIL (if there is only one homographic division)
1, 2, 3, n ...(otherwise)

in VanDale: NIL (if there is only one homographic division)
I, II, III, n ...(otherwise)

in LDOCE: NIL (if there is only one homographic division)
1, 2, 3, n ...(otherwise)

in OALD: NIL (if there is only one homographic division)
1, 2, 3, n ...(otherwise)
(or implicit/embedded)

in Vox: NIL (if there is only one homograph division)
1,2,3,n... (otherwise)

POS:

in Garzanti:

s.	=	noun
v.	=	verb
agg.	=	adjective

avv.	=	adverb
art.	=	article
prep.	=	preposition
pron.	=	pronoun
inter.	=	interjection
cong.	=	conjunction

in Collins:

Italian	English
s	n
v	v, vb
ag	adj
av	adv
art	art
prep	prep
pron	pron
escl	excl
cong	conj

in VanDale Monolingual:

1	=	noun
3	=	verb
2	=	adjective
5 or 2	=	adverb
75	=	article
6	=	preposition
4	=	pronoun
9	=	interjection
8	=	conjunction
70, 72	=	cardinal
73	=	ordinal

in VanDale Bilingual:
Dutch, the same as in the mono-lingual
English has no coding.

in LDOCE:

n
prep
interj
.....

in OALD:

n
prep
int(erjection)
.....

in Vox:

adj = adjective
adv = adverb
conj= conjuction
s = noun
v, vb = verb

prep = preposition
pron = pronoun
interj = interjection
loc = locution

subcat:

in Garzanti:
tr. = transitive, e.g. **presumere** *v.tr.*
ciancicare *v.tr.* e *intr.*
intr.= intransitive, e.g. **andare** *v.intr*
rifl = reflexive, e.g. **prestare** *v.tr....// -arsi v.rifl.*
arrendersi *v.rifl.*
arrampicare *v.intr...*, **arrampicarsi** *v.rifl.pron.*

in Collins:
t = transitive, e.g. **incorporate** *vt*
i = intransitive, e.g. **go** *vi*
r = reflexive, pronominal, reciprocal, e.g. **arrendersi** *vr*

in VanDale Monolingual and Bilingual:
for verbs : 1 intransitive
2 transitive
3 reflexive

in LDOCE: not present.

in OALD: t, i

in Vox:
impers, intr, prnl, tr, tr.-prnl,

subtype:

in Garzanti:
articolata = preposition combined with definite article,
e.g **del...***prep.articolata*
indef. = indefinite, e.g. **qualcosa...***pron.indef.*
interr. = interrogative, e.g. **quando...***avv.interr.*
num.card. = cardinal number, e.g. **quattrocento** . .
agg.num.card.
locuz. = locution, e.g. **ab ovo...***locuz.*
voce onom. = onomatopeic word, e.g. **ciac...***voce onom.*
invar. = invariable, e.g. **qualsivoglia...***agg.indef.invar.*
di tempo = of time, e.g. **dopodomani...***avv di tempo*
.....

in Collins:
impers = impersonal, e.g **be** *impers vb*
modal **will** *modal aux vb*
aux = auxiliary, **will** *modal aux vb*
.....

in VanDale Monolingual and Bilingual:

for pronouns:
1 personal pronoun
2 demonstrative pronoun
3 possessive ponoun
4 relative pronoun
5 interrogative pronoun
6 reflexive pronoun
7 reciprocal pronoun

for verbs :
4 auxiliary verb
5 copula
6 impersonal verbs like "rain", "snow"

in LDOCE:
various gram. labels
e.g. often pass
usu plural with singular meaning

in OALD:
tense, e.g. pp
pres p
aux
inf
imper
countable

in Vox:
(for adverb :) c (quantity)
m (manner)
l (place)
neg (negative)
o (order)
t (time)

(for noun:) pr (proper)

(for pronoun:) indef (indefinite)
relat (relative)

(for loc:) adj (adjectival)
conj (conjunctive)
adv (adverbial)
prep (prepositional)

gender:

in Garzanti:
f. e.g. **abilitazione** *s.f.*
m. **abete** *s.m.*

in Collins:
f **persecuzione** *sf*
m **sistema** *sm*

in VanDale Monolingual and Bilingual:

1	masculine
2	feminine
3	neutral
4	masc. and neut.
5	fem. and neut.
6	masc. and fem.
7	masc., fem. and neut.
8	neut. and 'male person'
9	neut. and 'female person'

in LDOCE: not present

in OALD:

masc = masculine
fem = feminine

in Vox:

f (feminine),
m (masculine),
com (common),
amb (ambiguous).

number:

in Garzanti:

sing.	e.g.	**io** *..pron.....sing.*
pl.		**forbici** *...s.f.pl.*

in Collins:

sg	**people** *sg*
pl	**informazioni** *fpl*

.....

in VanDale Monolingual:

mv.	plural
enk.	singular

in LDOCE: not present

in OALD: not present

in Vox:

sing
pl

g-code:

in VanDale: not explicitly present, perhaps extractable from the formalized examples.

in LDOCE: Cap No (Lett) + word qualifiers
e.g. **drink** *n* **1** [U;C] a liquid ...
dress up *v adv* ... **3** [T1 (as, in)]

in OALD: verb patterns (complementation + pred-arg structure)
No. (+ Cap)
e.g. **emerge** *vi* [VP2A, 3a] ...

Infl_label:

in Garzanti:

```
sing.  = singular
pl.    = plural
m.     = masculine
f.     = feminine
compar. = comparative
superl. = superlative
......
```

in Collins:

```
sg     = singular
pl     = plural
m      = masculine
f      = feminine
comp   = comparative
superl = superlative
pt     = past tense
pp     = past participle
...............
```

in VanDale Monolingual:

```
mv.       plural
enk.      singular
```

in LDOCE:

```
compar
superl
morph codes (types of suffix + idiosyncratic
             meaning with)
```

in OALD:

```
Comp(arative)
Superl(ative)
   (also found in <POS_subtype>)
```

in Vox:

sing	= singular
pl	= plural
f	= feminine
m	= masculine
superl	= superlative
aum	= augmentative
dim	= diminutive

Aux_label:

in Collins:

aus = auxiliary

in VanDale Monolingual:

h. auxiliary of perfect tense "hebben" ("have")
i. auxiliary of perfect tense "zijn" ("be")

in LDOCE:

Wv1, Wv2

in OALD: aux

Sense_No:

in Garzanti:

NIL (when there is only one word-sense within a HOM_GROUP)
e.g. **quadrettato** *agg.* suddiviso in quadrati.....

a number from 1 on (when there is more than one word-sense)
e.g. **quadrello**..*s.m.* 1 mattonelle....
2 [p.f.-a] *(lett.)* freccia....

in Collins:

NIL (if there is only one sense division)
(a), (b), (c), and so on (otherwise)

in VanDale:

a number from 1 on

in LDOCE:

nil (1)
number (>1)
subsenses - letters

in OALD:

nil (1) number/letter (>1)
(unnumbered secondary senses)

in Vox:

NIL (if there is only one sense division)
1,2,3,4,5, n... (otherwise)

Subject_code:

in Garzanti:
aer. = aeronautics
agr. = agriculture
bot. = botany
.....
e.g. **cima**.....4 *(bot)* infiorescenza....

in Collins:
Admin/Amm = Administration (for Eng-It and It-Eng sides)
Chem/Chim = Chemistry (" " " " ")
.....

in Van Dale:
cul. = culinair ("cooking")
astrol. = astrology
dipl. = diplomacy

in LDOCE:
two/four letter codes - subj / sub-subj

in OALD:
a few (e.g. accidence (gram))
(but not tagged and set provided is limited)

in Vox:

AERON. = aeronautics
ARQ. = architecture
ARQUEOL.= archaeology
ASTRON. = astronomy
ALBA. = masonry
ANAT. = anatomy
.....

Semantic_code:

in Garzanti:
fig. = figurative
.....

in Collins:
fig
e.g. **weak-kneed** *adj (fig)* debole,...

in VanDale:
fig. = figurative

in LDOCE:
fig
humor
euph

in OALD:
fig
hum
facet (tagged labels)

in Vox:

fig. = figurative
burl. = burlesque
desp./despec.= pejorative
eufem. = euphemism
iron. = ironic

Register_code:

in Garzanti:
scherz. = humorous
lett. = literary

in Collins:
frm = formal
fam = informal or colloquial
poet = literary or poetic usage

in VanDale:
euf. = euphemistic
iron. = ironic

in LDOCE:
infml
fml, etc.

in OALD:
archaic
colloq.
liter. (tagged labels)

in Vox:
lit. = literary
rust. = rustic
fam. = familiar
cientif.= scientific
fest. = homourous
pleb. = plebeian
poet. = poetic
vulg. = vulgar
neol. = neologism.

Usage_Code:

in Garzanti:
rar. = rare usage
e.g. **ibi** *s.m.invar.* (rar.) -> ibis

in Collins:
old = old fashioned
gen = generally, in most senses

in VanDale:
f = frequent

in LDOCE:
rare, etc.

in OALD:
mod use
old use (tagged labels)

In Vox:

ant. = old
desus. = not used
inus. = unusual
p.anal. = analogous
p.ant. =
p.excel.= preferably
p.ext. =
p.us. =
us. = used

Geographic_code

in Garzanti:
dial. = dialectal
region. = regional
.....

in Collins:
Scot = Scottish
Brit = British
Am = American

in VanDale Monolingual:
AZN = common in the south of the Netherlands and in Belgium
This is the only code used.

in VanDale Bilingual:
AE = American English
Sch.E. = Scottish English

in LDOCE:
AmE, AfrE, etc.

in OALD:
GB
F
Gk, etc.

in Vox:

dial.= dialectal
Al. = Alava
Albac.= Albacete
Alic. = Alicante
Alm. = Almeria
And. = Andalucia
Ar. = Aragon
Ast. = Asturias
Bad. = Badajoz
Burg. = Burgos
Cac. = Caceres
Cad. = Cadiz
Can. = Canarias
Cord. = Cordoba
C.Real. = Ciudad Real
Cuen. = Cuenca
Extr. = Extremadura
Gal. = Galicia
Gran. = Granada
Guadal.= Guadalajara
Guip. = Guipuzcoa
Logr. = LogroÙo
Mal. = Malaga
Murc. = Murcia
Nav. = Navarra
Pal. = Palencia
Sal. = Salamanca
Sant. = Santander
Seg. = Segovia
Sev. = Sevilla
Sor. = Soria
Tol. = Toledo
Val. = Valencia
Vallad.= Valladolid
Zam. = Zamora
Zar = Zaragoza

Country_code:

in Vox:

Amer. = American
Amer.Central.= Central America
Amer.Merid.= South America
Ant. = Antilles
Argent. = Argentine
Bol. = Bolivia
Colomb.= Colombia
C.Rica.= Costa Rica
Ecuad. = Ecuador
Filip. = Philippine

Guat. = Guatemala
Hond. = Honduras
Mej. = Mexico
Nicar.= Nicaragua
Pan. = Panama
Parag.= Paraguay
P.Rico.= Puerto Rico
R. de la Plata =Rio de la Plata
Salv. = El Salvador
S.Dom.= Santo Domingo (Dominican)
Urug.= Uruguay
Venez.= Venezuela

Semantic_Indicator_type:

in Collins:
the following types of semantic indicators can be derived: synonym
hypernym
contextual
typical_subject
typical_object
.....

in VanDale: similar types to Collins can be derived.

Taxon_label:

in Garzanti:
fam. = family
e.g. **mandorlo** ..... *(fam.Rosacee)*

in Collins:
no taxonomies.

in VanDale: not present

in LDOCE:
some box codes / some subj codes encode animacy, category membership, etc.

in OALD: not present

Trans_type:

in Collins:
explanation, e.g. **Befana** *sf* ... (b) *(personaggio)*
kind old woman who, according to legend, ...
cultural equivalent, e.g. **maturita'** "=" G.C.E. A-levels

in VanDale Bilingual the following types can be derived:
equivalent
near synonym
phrase
not equivalent
explanation

Trans_label:

in Collins:
"=" (see example above)

in VanDale Bilingual:
/[h93] means "not equivalent"
HVERT means "equivalent"
SYNVT means "near synonym of "HVERT"

Ex_Trans_type:

in Collins:
explanation, e.g. **liceo** *sm* ... liceo classico/ scientifico secondary school specializing in .....
cultural equivalent, e.g. **speaker** *n* ... (d) *(Brit Parliament):* the Speaker "=" il Presidente della Camera dei deputati.

Ex_Trans_label:

in Collins:
"="

Ex_label:

in Vox:
FR, fr, frs, EXPR,
Without any label, an example can appear elsewere in the definition introduced by a colon, followed by the text of the example written in italics and where the headword is sometimes referred as "~".

Prov_label:

in Garzanti:
prov. = proverb
e.g. **ciambella** .... 1..../*prov.*: non tutte le ciambelle riescono col buco, non sempre si riesce...2.

in Collins:
Proverb
e.g. **stitch** 1 *(Proverb)* un punto in tempo ne salva 100

in VanDale:
366 = a number reference to an external list of sayings which is in the book but not on the tape

in LDOCE: time label, geographic label

in OALD:
prov. label embedded in def in brackets - not tagged

Syn_label:

in Garzanti:
SIN.
e.g. **cima** 1. ..... SIN. **sommita', vetta,** .....

Collins does not list synonyms explicitly.

in VanDale:
SB

in LDOCE: compare .. as run-on

in OALD: not present

in Vox:
SIN.

Ant_label:

in Garzanti:
CONTR. = antonym
e.g. **chiuso** ... CONTR. **aperto**

Collins does not list antonyms explicitly.

in VanDale:
AB

in LDOCE:
opposite (occasionally as run on / xref)

in OALD:
opp(osite) label

in Vox:
CONTR. = antonym

Alt_label:

in Garzanti:
DIM. = diminutive
ACCR. = augmentative
...
e.g. casa .. 1...SIN...., DIM. **casina, casetta,** ACCR. **casona**

Collins does not list "altered forms" explicitly.

in VanDale: not present

in LDOCE: not present

in OALD: not present

in Vox:
aum. = augmentative
der. = derivative
dim. = diminutive
superl./SUPERL. = superlative,

Mwd_label:

in Garzanti: /

in Collins: cpd (on the Eng-It side)
: (on the It-Eng side)
e.g. **sewing** .. 1 *n* cucito 2 cpd: **sewing machine** *n*
macchina da cucire
guardia .. 1 *sf* ... 2: **guardia del corpo**
bodyguard

in VanDale: only indirectly derivable from the examples.

in LDOCE: various codes indicating type, syntax, stress, etc.

in OALD: nothing (<Pos_subtype> suggests deriv morph instead -
both tagged cd)

in Vox:

"~" and italics

Xref_type:

Values for our dictionaries: explicit
implicit
fig

Xref_label:

in Garzanti:
-> e.g. **henna** .. *s.f.* -> **henné** *s.m.invar* .....

in Collins:
= e.g. **pry** vt (Am) = **prise**
pt of **rode** *pt of* **ride**
pp of **ridden** *pp of* ride
abbr of TV *n* (abbr di television)
.....

in VanDale Monolingual:
vgl. +'aangebonden'

in LDOCE:
see also
compare, etc.

in OALD:
headword | compound/derivative

in Vox:
v./ V. = See
REL. = related.

Run-On_Type:

in Garzanti: adverbial suffix

in Collins: phrasal verb

in LDOCE: always derivational morphology (?)
full word, - part word + suffix or - suffix

Run-On_Label:

in Garzanti: //
e.g. **frontale** *agg.* ..../// **-mente** *avv.* di fronte

in Collins: "a solid lozenge" is used to indicate phrasal verbs at the end of an entry.

in VanDale: not present.

in LDOCE: not present

in OALD: not present

Homoph_label:

in Vox:
HOMOF.= equivalent sounds for different words.

SECTION 2

Definition of the Computational Model of the Common Dictionary Entry for the Project Lexical Database

2.1 Some General Observations on the Common Lexical Entry

When the analysis of the MRDs currently being used within the Project had been completed, we described them using a common representation language with the scope of making the data exchangeable and comparable between the Project partners while at the same time carefully maintaining and respecting the idiosyncracies of each single dictionary. After an evaluation of how the formalism had been applied to describe the different dictionaries, the next step was to design a Common Lexical Entry (CLE) into which each single Dictionary Entry could be mapped. The Common Lexical Entry has the following characteristics:
- it no longer reflects the linear order of the text of the single project dictionaries but can still contain all the information found in the source, even if this has been rearranged or is represented in a new way;
- it contains more types of "derived" information, i.e. information which is not explicitly present in the source dictionary, but which can be extracted with some processing;
- it is no longer bound and constrained by the necessity of compacting the data and saving space as in the printed medium;
- it can be easily expanded and updated.
Furthermore, the design of the Common Lexical Entry must be guided by the requirements of NLP, and criteria which are known to be useful in such systems must be respected.

The resulting Common Lexical Entry has been represented using the same formalism as that adopted for the individual dictionary templates. It can thus be described in the same terms used for the template description given in Section 1.3. It represents the "maximal" entry for the Common Lexical Database. The set of attributes adopted is for the most part the same as those used in Section 1; the CLE can, in fact, be considered to some extent as the "union" of the idiosyncratic entries. However, as the structure chosen for the CLE is different from that of the first templates and as much more "derived" information is now represented, some new Tags were necessary. Only these new tags will be described here, in section 2.3, following the CLE Template (for all the others, reference must be made to Section 1.3). Section 2.4. lists the possible values for these new tags.

With the specific templates for specific dictionaries of Section 1, we did not impose any standard hierarchy on the representation of the single dictionaries but simply described the lexical entries as they actually are while, at the same time, with the mapping of the information from each single dictionary onto the common tag system, we ensured uniformity in the interpretation of the contents of all the dictionaries. This first "lower" level of standardization imposes sufficient uniformity to be a useful basis for the exchange of MR dictionary documents. In the Template for the Common Entry, here in Section 2, we define a unique standard structure into which the data from any dictionary could be mapped. Conversion from the already explicit single dictionary templates to the CLE should be quite straightforward. The Common Lexical Entry Template has been designed to facilitate the organization of a common project lexical database in which not only the exchange of lexical data is involved but also data processing activities in collaboration are expected.

It has been decided to structure the entries in the CLE on a "one-entry one-major-part-of-speech" basis. This means that, where in the source dictionaries homographs representing more than one part of speech are compacted together in a single entry, they will now be split to create one DB entry for each separate part of speech while, however, maintaining a link between them. In the same way, much other information which was previously found in a single entry in many dictionaries, e.g. variant forms, derivatives formed from the headword by adding a suffix, phrasal verbs, etc., is represented in the DB in separate entries. It has thus been essential to represent explicitly relations between lexical items which were given implicitly in the source dictionary, but which risk being lost once the splitting of source entries has been completed. Thus, for each separate entry in the DB, the entry type is specified and all the entries which are indicated in the source dictionary as being related to this entry are listed in successive Related_Entry fields, each one followed by indication of the particular relationship involved in a Related_entry_type field; for example, the homographic relationship between DB entries with the same form but different major POSs, previously compacted under the same headword in the source dictionary, is identified by the value "Gram_Hom" entered in the Related_entry_type field. In the same way, values will be

entered in this field for all those entries which are shown in the source dictionary as being related in a dependent fashion to a given entry, e.g. variant forms, run-ons, etc.. The relationship between entries with the same graphical form (homographs) which appear in the source dictionary as separate entries has been maintained. The number which appeared in the source dictionary indicating the occurrence of more than one entry for the same form is entered as value in the Hdwd_Hom_No. field. A detailed explanation of all the Entry_ and Related_Entry_Group tags, the values which they can assume and examples of how particular entries are treated are given in Sections 2.3 and 2.4.

As we now have just one major POS for each entry, the POS tag has been extracted from the Gram_Inf_Group and has been raised out of the Hom_Group (now eliminated), appearing at the Headword_Group level. However, given that the Lexical Knowledge Base which is to be constructed by the Project will be used in NLP applications, which aim at working on a "word-sense" basis rather than at the word level, the entire Gram_Inf_Group has been "lowered" to the Sense_Group level, even though this will cause some redundancy of information stored. This decision was considered more theoretically sound and more efficient on the practical level. However, given that there are some kinds of information that apply to all senses and that might be relevant at the sense level, such as frequency information, grammatical homography, morphological variants, etc., which are currently stored at the headword level, it will be one of our goals in the Project to establish what is the balance between storage at the headword level and storage at the sense level. The LDB system should enable us to access all information regardless of the level (headword or sense) and it will only be after studying the dependencies of the pieces of information from all points of view that we can then decide which information is to be stored at which level in the final Knowledge Base.

Section 2.5 gives a brief description of the DB model which has been decided on for the project; eventually, the Computational Model of the Dictionary Entry for the Project LDB is constituted by the combination of the LE representation and the LDB model.

2.2 COMMON LEXICAL ENTRY TEMPLATE

Tags at Dictionary Level:

```
DICT_SOURCE
   Dict_name:
   Dict_type:

   LANGUAGE
M     Lang:
B       L1:
B       L2:
B     Metalanguage:

   PHONETIC TRANSCRIPTION
        IPA:
        Notes:
```

Tags at Entry Level:

```
 ENTRY_GROUP
   DB_Entry_Id.:
   Source_Entry_Id:
   Entry_type:

   HEADWORD_GROUP
    (Hdwd_text):
     Hdwd_form:
     Hdwd_type:
    (Hdwd_label):
    (Hdwd_Hom_No.):
    (Hdwd_freq_inf):
     Hdwd_POS:

     *(RELATED_ENTRY_GROUP)
         Related_entry:
        (Related_entry_extens):
         Related_entry_type:

      (PHONETIC_GROUP)
         (Pronunc_text):
         +Pronunciation:
         (Primary_stress_pos):
         (Secondary_stress_pos):

      (ETYMOLOGY_GROUP)
          Etymology_text:

     *(SENSE_GROUP)
          Sense_no.:
```

```
(CROSS_REFERENCE_GROUP)
   X-ref_type:
   Related_entry:
   Related_entry_extens:

(GRAM_INF_GROUP)
  (Gram_Inf_text):
  (Hom_form):
  (subcat):
  (subtype):
  (gender):
  (number):
  (various):
  (g_code):
  (aux_form):
  (INFLECTION_GROUP)
     (Infl_text):
     (Infl_label):
     (Infl_stem):
     +Infl_form:

(SENSE_LABEL_GROUP)

    (SEMANTIC_FEATURES_GROUP)
       (Sem_Features):
       (Sem_Roles):

    (SEMANTIC_LABEL_GROUP)
       (Semantic_label_text):
       (Subject_code):
       (Semantic_code):
       (Register_code):
       (Usage_code):
       (Geographic_code):
       (Country_code):

    (SEMANTIC_INDICATOR_GROUP)
       (Semantic_Ind_type):
        Semantic_Ind_text:

M    +DEFINITION_GROUP
       (Def_no):
       (CROSS_REFERENCE_GROUP)..
       (SEMANTIC_LABEL_GROUP)...
        Def_text:
          (TAXONOMY_GROUP)
             Taxon_label:
             Taxon_text:
```

```
B          +TRANSLATION_GROUP
              (Trans_no):
              (Trans_type):
              (SEMANTIC_LABEL_GROUP)...
              (SEMANTIC_INDICATOR_GROUP):...
               Trans_text:
              (GRAM_INF_GROUP)...
             *(TRANSLATION_GROUP)...

           *(EXAMPLE_GROUP)
               Ex_type:
              (SEMANTIC_LABEL_GROUP)...
              (SEMANTIC_INDICATOR_GROUP):...
               Example:
              (Ex_explanation):
              (COLLOCATION_GROUP)
                  Coll_POS:
                  Coll_word:

B             +EXAMPLE_TRANS_GROUP
                 (Ex_Trans_type):
                 (Ex_Trans_label):
                  Ex_Trans_text:
                 (GRAM_INF_GROUP)...
                 (SEMANTIC_LABEL_GROUP)...
                 (SEMANTIC_INDICATOR_GROUP):...

          *(MULTIWORD_GROUP)
               (Mwd_type):
               (SEMANTIC_LABEL_GROUP)...
                Mwd_form:
                Mwd_explanation:

              *(PROVERB_GROUP)
                   Prov_text:
                  (Prov_explan):

           (SEMANTIC_RELATIONS_GROUP)

              *(SUPERORDINATE_GROUP)
                   Genus_term:
                   Genus_term_extens.:

              *(SYNONYM_GROUP)
                   Synonym:
                   Syn_entry_extens:

              *(ANTONYM_GROUP)
                   Antonym:
                   Ant_entry_extens:

               (ALTERATE_GROUP)
                  +Alterate:
```

ABBREVIATED SCHEMA OF THE COMMON LEXICAL ENTRY

```
 ENTRY_GROUP
     HEADWORD_GROUP
       RELATED_ENTRY_GROUP
       PHONETIC_GROUP
       ETYMOLOGY_GROUP
       SENSE_GROUP
              CROSS_REFERENCE_GROUP
              GRAM_INF_GROUP
                 INFLECTION_GROUP
              SENSE_LABEL_GROUP
                 SEMANTIC_FEATURES_GROUP
                 SEMANTIC_LABEL_GROUP
                 SEMANTIC_INDICATOR_GROUP
M             DEFINITION_GROUP
                    TAXONOMY_GROUP
B             TRANSLATION_GROUP
              EXAMPLE_GROUP
B                   EXAMPLE_TRANS_GROUP
              MULTIWORD_GROUP
                    PROVERB_GROUP
              SEMANTIC_RELATIONS_GROUP
                     SUPERORDINATE_GROUP
                     SYNONYM_GROUP
                     ANTONYM_GROUP
                     ALTERATE_GROUP
```

2.3 Explanation of New Tags in the CLE

In this section we will only describe the nodes and tags in the CLE which are new or differ in some way from those used to represent the single project dictionaries.

Tags at Dictionary level:

Dict_type: we have changed the unstructured Dict_notes tag into a Dict_type tag which will be used to define the MRD data by a number of keywords chosen from a closed set. Possible values are listed in Section 2.4.

Tags at Entry Level:

ENTRY_GROUP = node grouping general information on the entry. In particular, this node specifies the type of entry being treated.

DB_entry_id: takes as value a number identifying uniquely the entry in the Project Lexical Database. It can be used as a pointer from other entries.

Source_entry_id: takes as value the number identifying the entry on the source tape so that any DB entry can be referred back to its source. DB entries which were compacted in the same dictionary entry can be recognized as they have the same value for this tag.

Entry_type: This tag indicates the particular type of entry. The possible values are: Main Entry, X-ref, Variant, Run-on, Phrasal verb, Homophone. They refer to the status of the entry in the source dictionary.

HEADWORD_GROUP = node grouping all general information on the headword.

Hdwd_text: This field is equivalent to the Hdwd_text field in the original templates and is used to preserve the source text of the headword in cases in which it contains hyphenation or stress information so that this is available for future analyses, as desired.

Hdwd_form: obligatory field, equivalent to the Hdwd_form field in the original templates; takes as value the dictionary citation form as it has been entered in this field in the source dictionary template, i.e. all signs which are additional to the actual graphic form of the headword have been removed, e.g.indications of hyphenation, stress, etc., whereas all signs which belong to the usual graphic form of the headword have been maintained, e.g. capitals, graphic accents, periods, spaces, etc.

Hdwd_type: tag indicating the particular type of headword. The possible values for each Entry_type are listed in Section 2.4.

Hdwd_label: this tag may be necessary when the Hdwd_type is a variant, and takes as value the label given in the source entry to specify the type of variant, e.g. rare usage, geographical variant, etc..

Hdwd_Hom_No.: takes as value any number appearing in the source dictionary, and therefore on the first-stage template, representing the occurrence of separate entries for the same graphical form. This field is necessary at this stage in order to represent relationships between entries which were present in the source dictionary.

Hdwd_POS: takes as values the major parts-of-speech (verb, noun, etc.) as they are represented in each dictionary. The list of possible labels for our dictionaries was given in Section 1.4 for the POS tag under the Gram_Inf_Group.
In the future,either a standard list or conversion tables for the individual lists of each source dictionary will be defined for the project.

RELATED_ENTRY_GROUP = node grouping information on the entries which were related to the entry in the source dictionary.

Related_entry: form of the related entry.

Related_entry_extens: takes as value the DB_entry_id of the related entry; this field should also contain a reference to the relevant Sense_no. of the related entry, when necessary.

Related_entry_type: this tag indicates the particular relationship between the entry and the related entry in the source dictionary.
The possible values are given in Section 2.4.

SEMANTIC_FEATURES_GROUP = the exact contents and format of this group will be decided later on.

SUPERORDINATE_GROUP = node grouping information on superordinates. If more than one superordinate is present in a definition, this node is repeated.

Genus_term: the superordinate which is extracted from the definition.

Genus_term_extens:takes as value the DB_entry_id of the genus term entry; this field should also contain a reference to the Sense_no., if necessary.

2.4 List of Values for the new Attribute Tags

The definitive list of values for the Common Lexical Entry will be given at the 30 month stage with the presentation of the final report on the Computational Model for the Lexical Entry. The values given here below refer only to the tags in the CLE which were not present in the earlier templates and are only indicative. In the final report, where possible, lists of standardized values will be given.

Dict_type:

Possible values are: Historical
Contemporary
Learner's
Thesaurus
Synonym
.....

Entry_type:

Main Entry
X-ref
Run-on
Variant
Phrasal Verb
Homophone

Hdwd_type:

The Hdwd_type is dependent on the Entry type.
Possible values for a Main Entry or a Variant are:

- lemma
- suffix
- prefix
- proper noun
- compound (e.g. bathing costume)
- phrasal verb
- letter
- .....

Possible values for a X-ref are:

- lemma
- word-form (inflected word-forms, usually irregular)
- abbreviation
- acronym
- contracted form (e.g. ain't)
- graphical variant
- .....

Possible values for a run-on are:

- derivative
- phrasal verb
- .....

Related_Entry_type:

The Related_entry_type is dependent on the Entry_type.

Possible values for the Related_entry_type of a Main Entry are:

Gram_Hom
Variant
Run-on
.....

Possible values for the Related_entry_type of a X-ref are:

Main Entry
Run-on
.....

Possible values for the Related_entry_type of a run-on are:

Main Entry
Variant
.....

Possible values for the Related_entry_type of a variant are:

Main Entry
Run-on
.....

Here below are examples of how different entries taken from OALD, 1974, will map into the CLE. We have included them here hoping that they make it easier to understand how the Entry, Headword and Related Entry Groups will function.

Bounteous

```
ENTRY_GROUP
  DB_Entry_Id.:nnn
  Source_Entry_Id:
  Entry_type:  ME

  HEADWORD_GROUP
    Hdwd_text:
    Hdwd_type:  lemma
    Hdwd_form:  bounteous
   (Hdwd_Homonym_No.):
    Hdwd_freq_inf:
    Hdwd_POS: adj

      RELATED_ENTRY_GROUP
        Related_entry:  bounteously
        Related_entry_extens.:
        Related_ent y_type:  Run_On
```

Bountiful

```
ENTRY_GROUP
  DB_Entry_Id.:
  Source_Entry_Id:
  Entry_type:  X_Reference

  HEADWORD_GROUP
    Hdwd_text:
    Hdwd_type: lemma
    Hdwd_form: bountiful
   (Hdwd_Homonym_No.):
    Hdwd_freq_inf:
    Hdwd_POS: adj

      RELATED_ENTRY_GROUP
       +Related_entry:  bounteous
        Related_entry_extens.:
        Related_entry_type:  ME

      RELATED_ENTRY_GROUP
       +Related_entry:  bountifully
        Related_entry_extens.:
        Related_entry_type:  Run_On
```

Bounteously

```
ENTRY_GROUP
  DB_Entry_Id.:
  Source_Entry_Id:
  Entry_type: Run_On

  HEADWORD_GROUP
    Hdwd_text:
    Hdwd_type:  lemma
    Hdwd_form:  bounteously
   (Hdwd_Homonym_No.):
    Hdwd_freq_inf:
    Hdwd_POS: adv

      RELATED_ENTRY_GROUP
       +Related_entry:  bounteous
        Related_entry_DB_Id: nnn
        Related_entry_extens.:
        Related_entry_type:  ME
```

Scoff1 (verb)

```
ENTRY_GROUP
  DB_Entry_Id.:
  Source_Entry_Id:
  Entry_type:  ME

  HEADWORD_GROUP
    Hdwd_text:
    Hdwd_type:  lemma
    Hdwd_form:  scoff
    Hdwd_Homonym_No.:  1
    Hdwd_freq_inf:
    Hdwd_POS: v

      RELATED_ENTRY_GROUP
       +Related_entry:  scoff
        Related_entry_extens.:
        Related_entry_type:  Grammatical Homograph

      RELATED_ENTRY_GROUP
       +Related_entry:  scoffer
        Related_entry_extens.:
        Related_entry_type:  Run_On

      RELATED_ENTRY_GROUP
       +Related_entry:  scoffingly
        Related_entry_extens.:
        Related_entry_type:  Run_On
```

Scoff1 (noun)

```
ENTRY_GROUP
  DB_Entry_Id.:
  Source_Entry_Id:
  Entry_type:  ME

  HEADWORD_GROUP
    Hdwd_text:
    Hdwd_type:  lemma
    Hdwd_form:  scoff
    Hdwd_Homonym_No.: 1
    Hdwd_freq_inf:
    Hdwd_POS: n

      RELATED_ENTRY_GROUP
       +Related_entry:  scoff
        Related_entry_extens.:
        Related_entry_type:  Grammatical Homograph
```

Sceptre

```
ENTRY_GROUP
  DB_Entry_Id.:
  Source_Entry_Id:
  Entry_type:  ME

  HEADWORD_GROUP
    Hdwd_text:
    Hdwd_type:  lemma
    Hdwd_form:  sceptre
   (Hdwd_Homonym_No.):
    Hdwd_freq_inf:
    Hdwd_POS: n

      RELATED_ENTRY_GROUP
       +Related_entry:  scepter
        Related_entry_extens.:
        Related_entry_type: Variant

      RELATED_ENTRY_GROUP
       +Related_entry:  sceptred
        Related_entry_extens.:
        Related_entry_type:  Run_On
```

Scepter

```
ENTRY_GROUP
  DB_Entry_Id.:
  Source_Entry_Id:
  Entry_type:  X_Reference

  HEADWORD_GROUP
    Hdwd_text:
    Hdwd_type:  lemma
    Hdwd_form:  scepter
   (Hdwd_Homonym_No.):
    Hdwd_freq_inf:
    Hdwd_POS: n

      RELATED_ENTRY_GROUP
       +Related_entry:  sceptre
        Related_entry_extens.:
        Related_entry_type:  ME
```

Sceptred

```
ENTRY_GROUP
  DB_Entry_Id.:
  Source_Entry_Id:
  Entry_type:  Run_On

  HEADWORD_GROUP
    Hdwd_text:
    Hdwd_type:  lemma
    Hdwd_form:  sceptred
   (Hdwd_Homonym_No.):
    Hdwd_POS: adj

      RELATED_ENTRY_GROUP
       +Related_entry:  sceptered
        Related_entry_extens.:
        Related_entry_type:  Variant

      RELATED_ENTRY_GROUP
       +Related_entry:  sceptre
        Related_entry_extens.:
        Related_entry_type:  ME
```

Sceptered

```
ENTRY_GROUP
  DB_Entry_Id.:
  Source_Entry_Id:
  Entry_type:  Variant

  HEADWORD_GROUP
    Hdwd_text:
    Hdwd_type:  lemma
    Hdwd_form:  sceptered
   (Hdwd_Homonym_No.):
    Hdwd_freq_inf:
    Hdwd_POS: adj

      RELATED_ENTRY_GROUP
       +Related_entry:  sceptred
        Related_entry_extens.:
        Related_entry_type:  Run_On
```

2.5 The Database Model

There is still no consensus on what would constitute a general computational model of a dictionary. Boguraev et al. (1990) have discussed the possible classes of dictionary models

> "We identify four classes of dictionary models. The first of these follows directly the well established notion of relational database, and tends to map a dictionary entry to a set of tables (see e.g. Fontenelle and Vanandroye, 1989). While this *relational* model of the lexicon can make use of a substantial body of research on database technology, it turns out to be the least suitable for mapping dictionaries to, given the intricate nature of lexical data.
>
> An improvement to this scheme is suggested by the insight that dictionary entries can quite naturally be regarded as shallow hierarchies with an open-ended number of attributes at each level (e.g.word sense clusters within a homograph or examples within a definition). A *hierarchical* model of the dictionary, then, offers advantages over the relational one in at least two respects (Neff *et al*, 1987). Firstly, it underlies a structured representation designed to encode the majority of existing conventions and notations for writing dictionary entries. Secondly, it suits lexical intuitions.
>
> While particularly well suited to most of the tasks of computational lexicology (see e.g. Boguraev, 1990), this model fails to meet at least one crucial requirement of computational lexicography. In particular, it offers no natural way of supporting an inverse transformation to that of parsing a dictionary source: the model is designed to encode complex structural relationships between fields and contents of a dictionary entry, but not at all to facilitate the derivation of a visual equivalent of these relationships (conventionally denoted by intricate typography).
>
> In contrast, and specifically for the purpose of incorporating a model of the lexicon into a (generic) lexicographer's workstation, a number of proposals elaborate the notion of a *tagged* dictionary representation (e.g. Amsler and Tompa, 1988). The tagged model places the emphasis on preserving all of the information associated with the form of a dictionary entry; however, it does not offer a natural way

of making explicit statements concerning structural relationships between its individual data fragments. Consequently, accessing the dictionary on the basis of structural, rather than notational, specifications is unintuitive, cumbersome, and sometimes impossible.

The two models are clearly complementary to each other. A generalisation of a particular technique, designed explicitly to support an open-ended range of requests from an on-line dictionary (Alshawi *et al*, 1989), aims at bringing the perspectives of the hierarchical and tagged models together. A *two-level* model of the lexicon retains an arbitrarily deeply (hierarchical) tagged isomorph of the source as the primary repository of lexical data ; at a separate level, a set of arbitrarily complex and interrelated indices implement any statement concerning the content and\or the form of the dictionary."

In our Project, this two-level "hierarchical-tagged" model has been chosen to represent dictionary entries in a form suitable for access and processing, while at the same time preserving an explicit representation of source data.

BIBLIOGRAPHY

Project Dictionaries

Il Nuovo Dizionario Italiano Garzanti, Garzanti, Milano, 1984.

Collins Concise English-Italian, Italian-English Dictionary, Collins, London and Glasgow, 1985.

Van Dale Groot Woordenboek Hedendaags Nederlands, P.G.J. van Sterkenburg and W.J.J. Pijnenburg (eds.), Van Dale Lexicografie, Utrecht/Antwerpen, 1984.

Van Dale Groot Woordenboek Nederlands-Engels, W.Martin and G.A.J.Tops (eds.), Van Dale Lexicografie, Utrecht/Antwerpen, 1986.

Longman Dictionary of Contemporary English (LDOCE), P. Procter et al. (eds.), Longman, Harlow and London, 1978.

Oxford Advanced Learner's Dictionary of Current English (OALD), A.S.Hornby (ed.)., Oxford University Press, Oxford, 1974.

Vox: Diccionario General Ilustrado de la Lengua Española, Bibliograf S.A., 1987.

References

Alshawi H.,Boguraev B. and Carter D. (1989), Placing LDOCE on-line, in B.Boguraev and E.Briscoe (eds.) *Computational Lexicography for Natural Language Processing*, Longman, London, 41-64.

Amsler R. and Tompa F. (1988), An SGML-based standard for English monolingual dictionaries, *Proceedings of the Fourth Annual Conference of the UW Center for the New OED: Information in Text*, Waterloo, 61-79.

Atkins B. (1989), Building a lexicon: beware of the dictionary, MS, Oxford University Press, presented at a BBN Symposium on Natural Language Processing, Cambridge, Mass.

Boguraev B. (1989), Building a lexicon: the contribution of computational lexicology, MS, IBM T.J.Watson Research Center, presented at a BBN Symposium on Natural Language Processing, Cambridge, Mass.

Boguraev B., Briscoe E, Carroll J., Copestake A. (1990) Database Models for Computational Lexicography, accepted for the EURALEX Conference, Malaga, Spain, 28-31 August 1990.

Fontenelle T. and Vanandroye J. (1989), Retrieving ergative verbs from a lexical database, MS, English Department, University of Liege.

Fought J., Van Ess-Dykema C., Toward an SGML Document Type Definition for Bilingual Dictionaries, TEI Internal Report, 1990.

Neff M., Byrd R. and Rizk O. (1987), Creating and querying hierarchical lexical data bases, in *Proceedings of the Second ACL Conference on Applied Natural Language Processing*, Austin, Texas, 84-93.

Neff M. and Boguraev B. (1989), Dictionaries, dictionary grammars and dictionary entry parsing, in *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, 91-101.

Picchi E., Peters C., Calzolari N., Implementing a Bilingual Lexical Database System, in *Proceedings of BUDALEX '88*, to appear.

TEI 1989, Text Encoding Initiative: Proposal for a Second Development Cycle, Technical Report TEI SCG 10, ACH, ACL, ALLC.