The Structure and Content of the Body of an OLIF v.2.0/2.1 File

Susan McCormick, OLIF Consortium
smccormick@comcast.net

1 General

The data in an OLIF v.2.0/2.1 (hereafter OLIF) file is organized in three main data groups:

1. The *header* contains data relevant to all of the lexical/terminological entries in the file.
2. The *body* contains the individual lexical/terminological entries.
3. *Shared resources* contains supplemental data (e.g., bibliographical information).

In this document, we present a description of the structure and content of the **body** of the OLIF file. For the formal description of the file header and shared resources, see current format documentation at www.olif.net.

2 The structure of the body of an OLIF file

The body of an OLIF file is a list of **entries** that contain data that is grouped according to the linguistic/lexical/terminological character of the information being represented. Since a primary motivator for OLIF was to offer a bridge between natural language processing lexicons (especially for MT) and terminology management applications, it has been designed with both the lexical and terminological view of the data in mind. The structure of an OLIF entry in the body of an OLIF file accordingly reflects a hybrid representation, with neither the explicit lemma-orientation of many lexicons, nor the explicit concept-orientation (with formal concept and term levels) of many terminology management models.

2.1 The three main data groups

To accommodate both the lexical and terminological models, OLIF developers have opted for a flexible structure based on **word-sense orientation**. For the specific purposes of describing OLIF, this means that **an OLIF entry is defined as a collection of monolingual data on a *specified sense of the word or phrase*, with optional links to represent transfer and cross-reference relations.** An OLIF entry accordingly has an obligatory grouping of monolingual data, and optional groupings of transfer and cross-reference data:

- **monolingual**: defines monolingual data; each OLIF entry may contain only one monolingual group.
- **cross-reference:** defines cross-reference relations between the given entry and other entries in the lexicon in the same language; while each cross-reference group in an OLIF entry represents a single cross-reference, there may be multiple cross-reference groups in the entry to represent multiple cross-references.
- **transfer:** defines transfer relations between the given entry and other entries in different languages; each transfer group in an OLIF entry represents a single, unidirectional transfer relation; multiple transfers (i.e., either to the same transfer language or to several different transfer languages) are represented by multiple transfer groups within the entry.

2.2 The key data categories

The OLIF word sense is itself defined as a semantic unit that is identified uniquely by a set of five **key data categories**:

- **canonical form:** the entry string, represented in canonical form in accordance with OLIF guidelines.
- **language:** the language represented by the entry string.
- **part of speech:** the part of speech, or word class, represented by the entry string.
- **subject field:** the knowledge domain to which the lexical/terminological entry is assigned.
- **semantic reading:** the semantic class identifier used to distinguish readings for entries with identical values for *canonical form, language, part of speech,* and *subject field.*

The key data categories together specify a given word sense and are required in the monolingual group of the entry in order to identify the entry itself. Since transfer and cross-reference relations imply links with other word senses, the key data categories are obligatory as well in any transfer or cross-reference data grouping (see section 2.5 for descriptions of shorthand identifiers for the list of key data categories for transfer and cross-reference). In both the transfer and cross-reference groups, the key data categories identify the word sense that is pointed to in the relation*.

Given the specification of an obligatory monolingual data group and optional transfer and cross-reference groups, a minimal well-formed OLIF entry contains a monolingual group with values for the key data categories *canonical form, language, part of speech, subject field,* and *semantic reading*, as illustrated in the following XML implementation of OLIF:

```
<entry>
   <mono>
      <keyDC>
         <canForm>table</canForm>
         <language>en</language>
         <ptOfSpeech>noun</ptOfSpeech>
         <subjField>general</subjField>
         <semReading>86</semReading>
      </keyDC>
   </mono>
</entry>
```

2.3 The OLIF mono

The monolingual data (mono) within an entry is grouped according to its linguistic/lexical/terminological nature. The groups themselves are sub-lists of data category/value pairs†. For example, a typical OLIF entry might encode information on the English noun *table* with data groupings like *key, administrative, morphological, syntactic,* and *semantic*:

* Note that for the cross-reference group, the data category *language* is not required since cross-reference relations are defined as intralingual links.

† The data category/value pairs are represented in XML as tags that reflect the element types, attributes, and values defined in the XML DTD/Schema

```
<entry>
   <mono>
      <keyDC>
         <canForm>table</canForm>
         <language>en</language>
         <ptOfSpeech>noun</ptOfSpeech>
         <subjField>general</subjField>
         <semReading>86</semReading>
      </keyDC>
      <monoDC>
         <monoAdmin>
            <originator>Weber</originator>
            <adminStatus>ver</adminStatus>
         </monoAdmin>
         <monoMorph>
            <inflection>like: book,books</inflection>
         </monoMorph>
         <monoSyn>
            <synType>cnt</synType>
            <synFrame>[gencomp-opt]</synFrame>
         </monoSyn>
         <monoSem>
            <definition>An arrangement of words, numbers, or signs or
            combinations of them, as in parallel columns, to exhibit a set of
            facts or relations in a definite, compact, and comprehensive
            form.</definition>
            <semType>inform</semType>
         </monoSem>
      </monoDC>
   </mono>
</entry>
```

The OLIF Mono

The key data categories identify the mono uniquely and include data on *canonical form, language, part of speech, subject field,* and *semantic reading*; the administrative data categories refer to information that can be used to organize or identify the mono administratively (e.g., *originator, administrative status, geographical usage*); the morphological data categories contain a morphological description of the monolingual string (e.g., *inflection, gender*); the syntactic data categories refer to the syntactic behavior associated with the mono (e.g., *syntactic type, syntactic frame*); and the semantic data categories represent information on the semantic level of analysis for the mono (e.g., *semantic type, natural gender*).

2.4 Transfer and cross-reference in OLIF

While the mono element refers to the status and behavior of the entry string, the *transfer* and *cross reference* elements describe links to other entries for the given mono; a transfer represents a link to an entry in another language, whereas a cross-reference is a link to an entry in the same language.

Transfer in OLIF is defined as bilingual and unidirectional: Each transfer group in an entry 1) refers to a single link between two entries in different languages, and 2) implies a transfer from the source (i.e., the entry described in the mono) to the target (i.e., the entry described in the transfer). An OLIF entry may contain an unspecified number of transfer elements, meaning that the lexicographer can define multiple transfers to the same language (e.g., English *source* -> French *target1*, French *target2*…), and/or multiple transfers into different languages (e.g., English *source* -> German *target*, French *target*, Spanish *target*…). Restrictions on the scope of a transfer (e.g., *source x is target y in context z*) are represented in the transfer element of OLIF by *transfer restrictions* (see section 3.1).

The semantics of cross-reference in OLIF also imply a directionality of the link from the originating entry to the entry that is being referred to. For example, an entry for English *table* with a cross-reference to English *row* via the cross-reference relation *has-meronym* means that *table* is a whole which has a part *row*. The entry for English *row* may have a corresponding cross-reference to *table* for the relation *has-holonym*, indicating that it is a part of the whole *table*:

```
<entry>
  <mono>
    <keyDC>
      <canForm>table</canForm>
      <language>en</language>
      <ptOfSpeech>noun</ptOfSpeech>
      <subjField>general</subjField>
      <semReading>86</semReading>
    </keyDC>
    <monoDC>
      <monoAdmin>
        <originator>Weber</originator>
        <adminStatus>ver</adminStatus>
      </monoAdmin>
      <monoMorph>
        <inflection>like: book,books</inflection>
      </monoMorph>
      <monoSyn>
        <synType>cnt</synType>
        <synFrame>[gencomp-opt]</synFrame>
      </monoSyn>
      <monoSem>
        <definition>An arrangement of words, numbers, or signs or
        combinations of them, as in parallel columns, to exhibit a set of
        facts or relations in a definite, compact, and comprehensive form.
        </definition>
        <semType>inform</semType>
      </monoSem>
    </monoDC>
  </mono>
    <crossRefer>
      <keyDC>
        <canForm>row</canForm>
        <language>en</language>
        <ptOfSpeech>noun</ptOfSpeech>
        <subjField>general</subjField>
        <semReading>69</semReading>
      </keyDC>
      <crLinkType>has-meronym</crLinkType>
    </crossRefer>
    <transfer>
      <keyDC>
        <canForm>Tabelle</canForm>
        <language>de</language>
        <ptOfSpeech>noun</ptOfSpeech>
        <subjField>general</subjField>
        <semReading>86</semReading>
      </keyDC>
    </transfer>

</entry>
```

OLIF entry with cross-reference and transfer

Since the specification of transfer and cross-reference relations is optional, a minimal well-formed OLIF entry includes a mono group with the key data categories, which, as noted above, together serve to identify the entry uniquely. Users may find minimally-specified OLIF entries a useful alternative to simple comma-separated formats or similar skeletal modelings of term entries. The relatively flat format of OLIF also means that basic entries are fairly easy to generate and read. Moreover, the optional morphological, syntactic and semantic OLIF data categories offer the user many choices for a more robust lexical/terminological description.

2.5 Numeric identifiers for key data categories in transfer and cross-reference

The reader will note in section 3.1 that OLIF also provides for a more compressed representation by specifying options for numeric identifiers for the mono or key data categories. Either of these ID types can be used in place of the list of five key data categories in any transfer or cross-reference component as a less repetitive way of identifying the mono that is being linked to. Using OLIF mono or key IDs allows for a more efficient representation of the entry for English *table*, for instance:

```
<entry>
  <mono MonoUserID=0651443876>
    <keyDC>
      <canForm>table</canForm>
      <language>en</language>
      <ptOfSpeech>noun</ptOfSpeech>
      <subjField>general</subjField>
      <semReading>86</semReading>
    </keyDC>
    <monoDC>
      <monoAdmin>
        <originator>Weber</originator>
        <adminStatus>ver</adminStatus>
      </monoAdmin>
      <monoMorph>
        <inflection>like: book,books</inflection>
      </monoMorph>
      <monoSyn>
        <synType>cnt</synType>
        <synFrame>[gencomp-opt]</synFrame>
      </monoSyn>
      <monoSem>
        <definition>An arrangement of words, numbers, or signs or
        combinations of them, as in parallel columns, to exhibit a set of
        facts or relations in a definite, compact, and comprehensive form.
        </definition>
        <semType>inform</semType>
      </monoSem>
    </monoDC>
  </mono>
  <crossRefer CrTarget=0591112687>
    <crLinkType>has-meronym</crLinkType>
  </crossRefer>
  <transfer TrTarget=0931445987>
  </transfer>
</entry>
```

IDs in cross-reference and transfer

The name of the mono ID attribute in the XML implementation above indicates that the identifier value is defined by the user. Universal identifiers for the *mono* and *key data category* groups are also specified for OLIF and allow the user maximal interchange possibilities by referring to system-independent identifiers of entry strings.

2.5 Concept-orientation and lemma-orientation

Version 2.0/2.1 of OLIF is designed to provide various views of the data. Whereas the OLIF prototype solely supported the core OLIF structure of a monolingual entry with a unidirectional transfer element, version 2.0/2.1 is expanded to allow the user to define a supraorganization of entries. In v.2.0 and 2.1, entries can be formally organized on a conceptual basis, as is done with many terminology representation models; in v. 2.1, word senses can be associated with specified lemmas as well. The concept-orientation supports a terminological or ontological organization, while the lemma-orientation supports a classic lexical organization.

The concept and lemma identifiers are associated with the top-level group *entry.* The concept IDs (user-defined or universal) can be used to organize entries as equivalent word senses associated with the same concepts rather than source word senses associated with transfers. The figure below illustrates how the standard OLIF entry for English *table* can be reorganized with a concept ID. Rather than a single entry for *table* with a transfer element for its German translation, there are two entries construed as equivalent via the concept ID:

```
<entry ConceptUserId="0731F16CCCD2D3119B4D">
   <mono>
      <keyDC>
         <canForm>table</canForm>
         <language>en</language>
         <ptOfSpeech>noun</ptOfSpeech>
         <subjField>general</subjField>
         <semReading>86</semReading>
      </keyDC>
      <monoDC>
            .....
      </monoDC>
   </mono>
</entry>

<entry ConceptUserId="0731F16CCCD2D3119B4D">
   <mono>
      <keyDC>
         <canForm>Tabelle</canForm>
         <language>de</language>
         <ptOfSpeech>noun</ptOfSpeech>
         <subjField>general</subjField>
         <semReading>86</semReading>
      </keyDC>
      <monoDC>
            .....
      </monoDC>
   </mono>
</entry
```

Using a concept ID

With the entries for *table* and *Tabelle* related by means of a common concept ID, a bidirectional equivalence is implied, unlike the source-target transfer direction of the standard OLIF model.

The lemma ID permits the user to organize OLIF word senses in a given language under a unifying lemma:

```
<entry LemmaUserId="712">
   <mono>
      <keyDC>
         <canForm>way</canForm>
         <language>en</language>
         <ptOfSpeech>noun</ptOfSpeech>
         <subjField>general</subjField>
         <semReading>54</semReading>
      </keyDC>
      <monoDC>
            .....
      </monoDC>
   </mono>
</entry>

<entry LemmaUserId="712">
   <mono>
      <keyDC>
         <canForm>way</canForm>
         <language>en</language>
         <ptOfSpeech>noun</ptOfSpeech>
         <subjField>general</subjField>
         <semReading>18</semReading>
      </keyDC>
      <monoDC>
            .....
      </monoDC>
   </mono>
</entry
```

Using a lemma ID

Note that the differentiating factor in the two entries above is the value for the *semantic reading*, i.e., there are two word senses for English *way*, related as readings of the same lemma.

3 The Content of OLIF Entries

Data categories and values for OLIF entries are referred to in the tables and descriptions that follow. Data category names are, where possible, coordinated with the names of ISO 12620 data categories, and generally follow those naming conventions.

3.1 Table of Data Categories

The data categories listed in the following table comprise the set of data categories available to the user for specifying an OLIF entry. The values associated with these data categories are described in Section 3.3 of this document. (Header data categories are described separately as part of the OLIF2 technical group's documentation.)

- Note: Within an OLIF entry, data category/value pairs may theoretically be listed in any order within the group tags that delimit them; this free ordering may or may not be supportable, depending on the technical representation selected.

Data category group	Data category name	Description
Basic: ***Obligatory***		The basic data categories are those data categories that are required for a minimal well-formed OLIF entry.
	<entry>	The **entry** data category delimits the OLIF entry. In addition, the following data categories may be optionally associated with the obligatory **entry** data category: **conceptUserId**: The **conceptUserId** data category gives a user-defined identifier of a concept **conceptUniversalId:** The **conceptUniversalId** data category gives a universal identifier (i.e., one which is unique not only in the user's environment, but worldwide) of a concept. **lemmaUserId**[‡]: The **lemmaUserId** data category gives a user-defined identifier of a lemma.
	<mono>	The **mono** data category groups the **monolingual** data within an entry. In addition, the following data categories may be optionally associated with the obligatory **mono** data category: **monoUserId**: The **monoUserId** data category gives a user-defined identifier of a grouping of monolingual data categories. **monoUniversalId**: The **monoUniversalId** data category gives a universal identifier (i.e., one which is unique not only in the user's environment, but worldwide) of a grouping of monolingual data categories.

[‡] Version 2.1 only

	<keyDC>	The **key data category** designator groups the five key data categories whose values uniquely identify an OLIF entry: **canForm, language, ptOfSpeech, subjField,** and **semReading.** In addition, the following data categories may be optionally associated with the obligatory **keyDC**: **keyDCUserId:** The **keyDCUserId** data category gives a user-defined identifier of a grouping of OLIF key data categories. **keyDCUniversalId:** The **keyDCUniversalId** data category gives a universal identifier (i.e., one which is unique not only in the user's environment, but worldwide) of a grouping of OLIF key data categories.
	<canForm>	The **canonical form** designates the entry string, represented in canonical form, as specified in OLIF guidelines. In addition, the following data category is associated with the **canonical form** designator**:** **xml:lang**: The **xml:lang** data category indicates the language of the entry string; Used in addition to the **language** data category, it facilitates exchange with standards that also use **xml:lang**.
	<language>	Indicates the **language** to which the entry string belongs.
	<ptOfSpeech>	Indicates the **part of speech** represented by the entry string. (In cases of phrases/multiword entries, the value for part of speech depends on the function of the phrase/multiword within a clause; the part of speech of the head element often indicates the part of speech value for the entire phrase/multiword string.)
	<subjField>	The **subject field** refers to the knowledge domain to which the lexical/terminological entry is assigned.
	<semReading>	The **semantic reading** indicates the semantic class identifier used to distinguish readings for entries with identical values for *canonical form, language, part of speech,* and *subject field.*
General: ***Optional***	**<generalDC>**	The **general data category** designator groups the general data categories. General data categories are optional data categories that can be used in any of the OLIF groups (*mono, cross-reference,* or *transfer)*
	<updater>	The **updater** is the individual who last modified the entry.
	<modDate>	The **modification date** indicates the date that the entry was last modified.
	<example>	The **example** is a sample text or portion of text that contains the entry string as an illustration of usage.
	<usage>	Indicates a **usage** note for the entry string
	<note>	Refers to a **note**, or commentary, on an entry by the lexicographer/terminologist.

		In addition, the following optional data category may be associated with the **note** data category: **noteType:** The **noteType** data category can be used to categorize notes (e.g. 'for localizer', 'for quality management).
Monolingual: ***Optional***	**<monoDC>**	The **monolingual data category** designator groups the optional data categories that may be used only within the *mono* group: **monoAdmin, monoMorph, monoSyn,** and **monoSem.**
administrative:	**<monoAdmin>**	The **monolingual administrative** designator groups the administrative data categories within a monolingual entry.
	<userDesignat>	Indicates the **user designator** of the entry string; used if the obligatory canonical form does not closely resemble the surface form.
	<syllabification>	Indicates **syllable** boundaries within the entry string.
	<geogUsage>	Refers to the **geographical usage**, or dialect, to which the entry string belongs.
	<entryType>	The **entry type** refers to the status of the entry string as representing a *product name, trademark,* or *orthographic variant*.
	<entryFormation>	The **entry formation** indicates the shape/structure of the entry string.
	<phraseType>	Further specifies the **type of phrasal** entry string.
	<entryStatus>	Indicates the **status** of an **entry** within a given lexicon/termbase.
	<entrySource>	Refers to the **entry source**, or the lexicon/termbase that the entry originated from.
	<originator>	The **originator** is the individual who originated the entry.
	<adminStatus>	Indicates the **administrative status** of an entry relative to a given work environment
	<company>	Indicates the **company**/organisation for whom entry is valid.
	<abbrev>	Indicates an **abbreviated** form of the entry string.
	<orthVariant>	Indicates an **orthographic variant** for the entry string. In addition, the following optional data categories may be associated with the **orthVariant** data category: **varType:** The **varType** data category can be used to specify types of orthographic variants, *spelling, transcription.* **transSystem:** The **transSystem** data category is used to note the type of transcription used.
	<depSynonym>	Indicates a rejected or **deprecated synonym** for the entry string.
	<timeRestrict>	Refers to a **time restriction**, or the period of time during or since which usage of the entry is valid.
	<product>	Indicates a **product** for which the entry is valid.
	<project>	Indicates a **project** for which the entry is valid.

	<locInfo>	Refers to **localization**-relevant **information** (e.g., product version, component name, operating system platform, or build number).
	<confidence>	Indicates how **confident** a term extraction program is that a term really is a term.
morphological:	**<monoMorph>**	The **monolingual morphological** designator groups the morphological data categories within a monolingual entry.
	<morphStruct>	Provides a transcription of the **morphological structure** of the entry string.
	<inflection>	Encodes the **inflection** pattern(s) of the entry word or inflected element of multiword/phrasal entry.
	<head>	Indicates the **head** word in a multiword/phrasal entry string.
	<gender>	Indicates grammatical **gender.**.
	<case>	Indicates grammatical **case** designation.
	<number>	Indicates grammatical **number**.
	<person>	Indicates **person**.
	<tense>	Indicates verb **tense**.
	<mood>	Indicates **mood** or mode.
	<aspect>	Indicates verbal **aspect**.
	<degree>	Indicates adjectival **degree** type.
	<auxType>	Indicates the **auxiliary type** for an auxiliary verb.
syntactic:	**<monoSyn>**	The **monolingual syntactic** designator groups the syntactic data categories within a monolingual entry.
	<synType>	The **syntactic type** describes the general syntactic behavior of the entry string.
	<synPosition>	The **syntactic position** describes the unmarked positioning of the entry string syntactically.
	<transType>	Describes the **transitivity type** of a verb.
	<synStruct>	Indicates the constituent **structure** of a multiword entry string.
	<synFrame>	Describes the **syntactic frame** data categories for the entry string (subcategorisation).
	<prep>	**Preposition**; used to further specify syntactic frame data categories.
	<verbPart>	**Verb particle**; used to further specify syntactic frame data categories.
semantic:	**<monoSem>**	The **monolingual semantic** designator groups the semantic data categories within a monolingual entry.
	<definition>	The **definition** is a prose definition of the entry string.
	<natGender>	The **natural gender** refers to the biological gender associated with the entry.
	<semType>	The **semantic type** represents the status of the entry string with respect to a semantic type classification structure.
Cross-Reference:	**<crossRefer>**	The **cross-reference** designator defines cross-

Optional		reference relations between the given entry and other entries in the lexicon in the same language. It groups the cross-reference data within a monolingual entry. Within each **cross-reference** element, the **keyDC** data categories are obligatory. The obligatory **keyDC** data categories may be alternately represented in **cross-reference** by the following associated data category: **crTarget**: The crTarget identifier specifies the target entry of a cross-reference relationship.
	<crLinkType>	Indicates the **type** of **cross-reference link** that pertains between the entry from which the link originates and the entry to which the link points.
	<orthVariantType>	The **orthographic variant type** holds information about the type of orthographic variant that the target of a cross-reference represents.
Transfer: ***Optional***	**<transfer>**	The **transfer** data category defines bilingual transfer relations between the given entry and other entries in the lexicon in different languages. The transfer data category groups the transfer data within a monolingual entry. Within each transfer data category, the **keyDC** categories are obligatory. The obligatory **keyDC** data categories may be alternately represented in **transfer** by the following associated data category: **trTarget:** The **trTarget** data category specifies the target entry of a transfer relationship. In addition, the following optional data category may be associated with **transfer**: **trDefault:** The **trDefault** data category specifies whether the given transfer is the default transfer.
	<equival>	Encodes the degree of transfer relationship, or **equivalence**, between words/phrases in two different languages.
	<trRestrictStmt>	The **transfer restriction statement** is a container for grouping multiple related transfer restrictions.
	<trRestrict>	Expresses a single **transfer restriction**.
	<contextStmt>	The **context statement** is a logical expression about the context(s) specified in the transfer restriction or structural change.
	<context>	Indicates one of the following: 1) the **context** for a given translation of a source word/phrase into a target word/phrase, or 2) the **context** for a structural change in the target language.
	<logOp>	Designates a **logical operator**. Valid values are: AND, OR, and NOT for **trRestrictStmt** and AND for **structChangeStmt.**
	<testStmt>	The **test statement** states one or more tests on the context(s).
	<test>	States a single **test**.

	<testType>	Indicates the **type** of **test**. Valid values are: *string* and *datacat.*
	<testDC>	The **test data category** names the data category to which a test pertains.
	<testValue>	Describes the **value** of the string or data category being **tested** for the context(s).
	<structChangeStmt>	The **structural change statement** is a container for grouping multiple, related structural changes.
	<structChange>	Describes a **structural change** in the target language vis-à-vis the source structure based on a transfer restriction having been satisfied.
	<changeType>	Indicates the **type of change**, e.g., *addInTarget, delIntarget, changeRole, assignCase,* etc.
	<changePOS>	Names the **part of speech** of an element being added or deleted.
	<changeValue>	Describes the **value** of the string or data category being **changed**.

3.2 Values

3.2.1 Values for KEY Data categories

⇒ **All KEY data categories occur <u>obligatorily</u> in an entry in the monolingual group; they are also required within the cross-reference and/or transfer groups, if these groups are contained in the entry.**

(Please note the exception of the *language* data category in the cross-reference group.)

Canonical Form <canForm>

⇒ Entry string in canonical form
⇒ Value: string

The shape of the canonical form is based on language-specific guidelines issued by the OLIF2 consortium in cooperation with the SALT project.

Language <language>

⇒ Language represented by entry string
⇒ Value: any valid designator from ISO 639 1

Part of Speech <ptOfSpeech>

⇒ Part of speech of entry string
⇒ Values:

VALUE	DESCRIPTION
noun	noun
verb	verb
adj	adjective
adv	adverb
prep	preposition
conj	conjunction
det	determiner
part	verb particle
auxverb	auxiliary verb
pron	pronoun
punc	punctuation
other	other pos to be determined by user

Subject Field <subjField>

⇒ Knowledge domain to which lexical/terminological entry is assigned.
⇒ Values: basic values as follows (from Eurodicautom); user has option to expand to accommodate individual hierarchies

VALUE	DESCRIPTION
agriculture	farming and agriculture
audiovisual	audiovisual
aviation	aviation and aerospace
botany/zoology	botany and zoology
budget	budgets and accounting
chemistry	chemistry
construction	construction and building
customs	customs, duties
defense	defense
development	development
economics	economics
education	education
electrotechnics	electronics
employment	human resources, employment
energy	energy
environment	environment
eurospeak	common European language terminology
finance	finance
fisheries	fishery science and technology
general	general vocabulary
geology	geology
industry	industry and industrial policy
informatics	information technology, programming
insurance	insurance
law	law
mechanics	mechanics
medicine	medicine
mining	mining
nuclear	nuclear power, nuclear industry
social	social science and policy
statistics	statistics
steel	steel
taxation	taxes
technology	general technology
telecom	telecommunications
trade	trade and tariffs
transport	transportation

Semantic Reading <semReading>

⇒ Identifier used to distinguish readings for entries with identical values for *canonical form, language, part of speech,* and *subject field*
⇒ Values: several possibilities/issues have been discussed:
- The requirement of a semantic reading that actually reflects a lexical semantic analysis has the potential for inhibiting data exchange rather than facilitating it,

e.g., different users interpret the semantic class hierarchies differently, or, since they don't pay attention to these differences at all in their lexical data (e.g., they have only a few cases where they require a distinction & thus have most of their entries with no semantic reading designation), must make these judgments for the purpose of OLIF only.

- Numeric semantic identifier assigned by the user has the same problem that a reading no.has in terms of its meaning possibly not being valid outside of the particular data set
- Some suggestions:
  - Have a pre-ordained set of values (e.g., from SIMPLE), but also allow a value of 'unspecified' for the masses of entries for which there is only one reading – allowing users an opt-out from making these judgments for each entry.
  - As an option, allow the user to use numeric identifiers from an authority (specified in the header) for the given language.
  - Do not use the semantic reading as part of the primary key at all, but rather as a 'backup' secondary key, to be used for disambiguation purposes only.

- **As of April 2001: Consensus that a standard for each language could be selected, e.g., *Roget's* and the numbering scheme for word senses from the designated standard utilized.**

3.2.2 Values for GENERAL Data categories

⇒ **General data categories are optional data categories that can be used in any of the groups (*monolingual, cross-reference,* or *transfer*).**

Updater <updater>

⇒ Refers to individual who last modified entry
⇒ Value: string

Modification date <modDate>

⇒ Date entry was last modified
⇒ Value: date

Example <example>

⇒ Sample text or portion of text in which entry string occurs
⇒ Value: string

Usage Note <usage>

⇒ Open field for notes on usage of entry string
⇒ Value: string

Note <note>

⇒ Open field for commentary by lexicographers/terminologists
⇒ Value: string

3.2.3 Values for Optional MONOLINGUAL Data categories

⇒ **The following data categories are optional within the monolingual group.**

3.2.3.1 Administrative MONOLINGUAL Data categories

User Designation <userDesignat>

⇒ Indicates entry string in a more 'user-friendly' way if the obligatory canonical form does not closely resemble the surface form.
⇒ Values: string

Syllabification <syllabification>

⇒ Indicates syllable boundaries within entry string.
⇒ Values: string formulated based on following guideline:
- a syllable boundary is designated by the presence of the '-' character placed between the two characters where the boundary occurs, e.g., *can-dle*

Geographical Usage <geogUsage>

⇒ Dialect represented by entry string
⇒ Value: any valid designator as specified in ISO 12620 (A.2.3.2) using ISO 3166 (Represent combined language-country codes, e.g., de-CH, en-GB)

Entry Type <entryType>

⇒ Refers to the status of the entry string as a product name, trademark, orthographic variant
⇒ Values: as follows

VALUE	DESCRIPTION
product-name	product name
trademark	trademark
orth-var	orth-var
un	unspecified

Entry Formation <entryFormation>

⇒ Indicates shape/structure of entry string

⇒ Values: as follows

VALUE	DESCRIPTION
abb	abbreviation
acr	acronym
sgl	single word
cmp	compound
phr	phrase
un	unspecified

Phrase Type <phraseType> -

⇒ Further specifies the phrasal entry string
⇒ Values: as follows

VALUE	DESCRIPTION
mw	multiword
set-phr	fixed, lexicalized phrase
coll	collocation
idiom	idiom
un	unspecified

Entry Status <entryStatus>

⇒ Indicates status of entry within given lexicon/termbase
⇒ Values: as follows:

VALUE	DESCRIPTION
word	general vocabulary item
term	specific to non-general domain
concept	concept
stopword	stopword
un	unspecified

Entry Source <entrySource>

⇒ Indicates lexicon/termbase that entry originated from
⇒ Value: string

Originator <originator>

⇒ Refers to individual who created entry
⇒ Value: string

Administrative status <adminStatus>

⇒ Indicates administrative status of an entry relative to a given work environment
⇒ Values: as follows

VALUE	DESCRIPTION
new	new entry
ver	verified
def	defaulted
mt	for MT only
obs	obsolete
un	unspecified

Company <company>

⇒ Indicates company/organisation for whom entry is valid
⇒ Value: string

Abbreviation <abbrev>

⇒ Abbreviated form of entry string (alternative to cross-reference representation)
⇒ Value: string

Orthographic Variant <orthVariant>

⇒ Indicates orthographic variant for entry string (alternative to cross-reference representation)
⇒ Value: string

Deprecated Synonym <depSynonym>

⇒ Indicates rejected synonym for entry string
⇒ Value: string

Time Restriction <timeRestrict>

⇒ Indicates period of time during or since which usage of entry is valid
⇒ Value: string

Product <product>

⇒ Identifies product for which entry is valid
⇒ Value: string

Project <project>

⇒ Identifies project for which entry is valid
⇒ Value: string

Localisation Information <locInfo>

⇒ Refers to localization-relevant information (e.g., product version, component name, operating system platform, or build number).
⇒ Value: string

Confidence <confidence>

⇒ Used with term extraction; the value of the data category indicates how confident the term extraction program is that the term really is a term.
⇒ Value: string

3.2.3.2 Morphological MONOLINGUAL Data categories

Morphological Structure <morphStruct>

⇒ Provides a transcription of the morphological structure of the entry string
⇒ Value: the value is formulated based on the following guidelines:
- '#' designates a word boundary
- '+' designates boundary between affix-root or affix-affix
- ':' designates boundary between elements of a compound
- '[]' designates nested constituents

Inflection <inflection>

⇒ Encodes the language-specific inflection pattern(s) of the entry word or head of multiword/phrase entry.
⇒ Value: two value types possible:
1. 'Inflects like' value (provided by Logos for all languages)
2. User-specified schema (e.g., use of Wahrig numbered patterns for German)
⇒ Values for 'inflects-like' patterns for English, German, French, Spanish and Portuguese are available on the OLIF2 web site www.olif.net.

Head Word <head>

⇒ Indicates the head word in a multiword/phrasal entry string.
Value: string (representing the actual head word)

Gender <gender>

⇒ Indicates grammatical gender.
⇒ Value: as follows:

VALUE	DESCRIPTION
m	masculine
f	feminine
n	neuter
c	common
un	unspecified

Case <case>

⇒ Indicates case designation.
⇒ Value: as follows:

VALUE	DESCRIPTION
n	nominative
g	genitive
d	dative
a	accusative
obj	objective
subj	subjective
loc	locative
prp	prepositional
inst	instrumental
un	unspecified

Number <number>

⇒ Indicates number.
⇒ Value: as follows:

VALUE	DESCRIPTION
sg	singular
pl	plural
sgt	singularetantum
plt	pluraletantum
du	dual
invar	invariant
un	unspecified

Person <person>

⇒ Indicates person.
⇒ Value: as follows:

VALUE	DESCRIPTION
first	first person
sec	second person
third	third person
un	unspecified

Tense <tense>

⇒ Indicates verb tense.
⇒ Value: as follows:

VALUE	DESCRIPTION
pres	present
past	past
fut	future
un	unspecified

Mood <mood>

⇒ Indicates mood (or mode).
⇒ Value: as follows:

VALUE	DESCRIPTION
indic	indicative
subj	subjunctive
imper	imperative
cond	conditional
sup	supine
un	unspecified

Aspect <aspect>

⇒ Indicates verbal aspect.
⇒ Value: as follows:

VALUE	DESCRIPTION
simp	simple
perf	perfective
imperf	imperfective
dur	durative
habit	habitual
iter	iterative
un	unspecified

Degree Type <degree>

⇒ Indicates degree type for adjective.
⇒ Value: as follows:

VALUE	DESCRIPTION
pos	positive
comp	comparative
sup	superlative
ela	elative
un	unspecified

Auxiliary Type <auxType>

⇒ Indicates type of auxiliary verb.
⇒ Value: as follows:

VALUE	LANGUAGE DESCRIPTION
have	da
være	da
have	en
be	en
être	fr
avoir	fr
laisser	fr
faire	fr
haben	de
sein	de
werden	de
lassen	de
ter	pt
estar	pt
estar	es
haber	es
un	unspecified

3.2.3.3 Syntactic MONOLINGUAL Data categories

Syntactic Type <synType>

⇒ Describes the general syntactic behavior of the entry string.
⇒ Value: as follows:

PART OF SPEECH	VALUE	DESCRIPTION
Noun	cnt	countable noun
	mass	mass noun
	mass-cnt	countable mass noun
	prop	proper noun
	coll	collective noun
	quant	quantitative noun
	def	definite noun
	indef	indefinite noun
Verb	recip	reciprocal verb
	refl	reflexive verb
	aux	auxiliary verb
	main-vb	main verb
	modal	modal verb
Adjective	attrib	attributive adjective
	pred	predicative adjective
	poss-adj	possessive adjective
	able-adj	-able participle
	ppart	past participle
	prespart	present participle
Adverb	degree	indicates degree, e.g., 'too'
	adv-mod	modifies adverb
	adj-mod	modifies adjective
	cls-mod	modifies clause
	np-mod	modifies noun phrase
	nu-mod	modifies numeral
	prep-mod	modifies preposition
	det-mod	modifies determiner
	quant-mod	modifies quantifier
Preposition	loc	locative preposition
	dir	directional preposition
	temp	temporal preposition
Conjunction	conj	conjunction
	comp-conj	comparative conjunction
	subj-conj	subjunction
Determiner	def-det	definite determiner
	indef-det	indefinite determiner
	interr-det	interrogative determiner
	poss-det	possessive determiner
	rel-det	relative determiner
	demonst-det	demonstrative determiner
	quant-det	quantitative determiner
	part-det	partitive determiner
Pronoun	def-pro	definite pronoun
	indef-pro	indefinite pronoun
	interr-pro	interrogative pronoun
	poss-pro	possessive pronoun

	rel-pro	relative pronoun
	demonst-pro	demonstrative pronoun
	quant-pro	quantitative pronoun
	pers-pro	personal pronoun
	part-pro	partitive pronoun
	refl-pro	reflexive pronoun
	wh-pro	Wh-type pronoun
	un	unspecified

Syntactic Position <synPosition>

⇒ Describes the unmarked positioning of the entry string syntactically.
⇒ Value: as follows:

PART OF SPEECH	VALUE	DESCRIPTION
Adjective	prenoun	before noun
	postnoun	following noun
Adverb	preverb	before main verb
	postverb	following main verb
	cl-init	clause-initial
	cl-final	clause-final
	deg-post	degree adverb after morpheme
	deg-pre	degree adverb before morpheme
Preposition	prep	prepositional to noun head
	postp	postpositional to noun head
	circumprep	preposition in circum position
	circumpostp	postposition in circum position
	un	unspecified

Transitivity Type <transType>

⇒ Describes the transitivity behaviour of verbs and deverbal nouns
⇒ Value: as follows:

PART OF SPEECH	VALUE	DESCRIPTION
Verb, Deverbal Noun	trans	transitive
	intr	intransitive
	ditrans	ditransitive
	refl	reflexive
	mid	middle
	caus	causative
	unacc	unaccusative intransitive
	unerg	unergative intransitive
	un	unspecified

Syntactic Structure <synStruct>

⇒ Indicates the constituent structure of a multiword entry string.
⇒ Value: pending

Syntactic Frame <synFrame>

⇒ The syntactic frame describes the subcategorisation of the entry word/phrase. The approach taken here adapts and expands on the original OLIF analysis, which was essentially a slot-grammar approach. The lexicographer builds the frame by specifying individual frame data categories from the slot values table below. (Slot fillers are implied with many of the slot values, but are language-specific and not formally represented).

The syntax for the frame specifies the following conventions:

- the syntactic frame is enclosed in square ([ ]) brackets
- slots are separated by commas (,)
- slots that are or'ed together are enclosed in parenthees and separated by vertical slashes, e.g., (.|.|.|.)

Example of a possible syntactic frame for the English verb *try*:

[subj, (dobj-opt | dobj-sent-ing-opt | dobj-sent-inf-opt)]

(Note: Specific prepositions or particles that fill a pp or part slot are specifiable with the data categories *prep* and *part* (description follows).)

⇒ Value: as follows:

PART OF SPEECH	VALUE	DESCRIPTION
Verb	subj	subject NP required
	subj-sent-opt	sentential subject optional (e.g., finite clause, infinitive clause, -ing clause, wh-, finite with 'that', 'dass')
	subj-imps-opt	impersonal subject optional (e.g., "It is raining")
	dobj	direct object NP required
	dobj-opt	direct object NP optional
	dobj-sent-opt	sentential direct object optional (e.g., finite clause, infinitive clause, -ing clause, wh-, finite with 'that', 'dass')
	dobj-sent-fin-opt	finite clause direct object optional
	dobj-sent-inf-opt	infinitive clause direct object optional
	dobj-sent-ing-opt	-ing clause direct object optional
	dobj-sent-that-opt	that/dass-clause direct object optional
	dobj-sent-wh-opt	wh-clause direct object optional
	dobj-comp-opt	e.g., "They elected him *president*"
	iobj	indirect object NP required
	iobj-opt	indirect object NP optional
	iobj-sent-opt	sentential indirect object optional
	genobj	genitive object required
	genobj-opt	genitive object optional
	pred-opt	predicate nominal (incl.sentential)/predicate adj.

PART OF SPEECH	VALUE	DESCRIPTION
		optional
	vcomp-opt	sentential verb complement optional (e.g., finite clause, infinitive clause, -ing clause, wh-, finite with 'that', 'dass')
	vcomp-fin-opt	finite clause verb complement optional
	vcomp-inf-opt	infinitive clause verb complement optional
	vcomp-ing-opt	-ing clause verb complement optional
	vcomp-that-opt	that/dass-clause verb complement optional
	vcomp-wh-opt	wh-clause verb complement optional
	part	verb particle required
	part-opt	verb particle optional
Noun	gencomp-opt	Genitive phrase optional (e.g., "the book of John", "the reading of the will"
	ncomp-opt	sentential noun complement optional (e.g., finite clause, infinitive clause, -ing clause, wh-, finite with 'that')
	ncomp-fin-opt	finite clause noun complement optional
	ncomp-inf-opt	infinitive clause noun complement optional
	ncomp-ing-opt	-ing clause noun complement optional
	ncomp-that-opt	that-type clause noun complement optional
	ncomp-wh-opt	wh-clause noun complement optional
Adjective	adjcomp-opt	sentential adj complement optional (e.g., finite clause, infinitive clause, -ing clause, wh-, finite with 'that')
	adjcomp-fin-opt	finite clause adj complement optional
	adjcomp-inf-opt	infinitive clause adj complement optional
	adjcomp-ing-opt	-ing clause adj complement optional
	adjcomp-that-opt	that-type clause adj complement optional
	ncomp-wh-opt	wh-clause adj complement optional
Noun, Verb, Adjective		
	pp	prepositional phrase required
	pp-opt	prepositional phrase optional
	pp-loc	locational/directional prepositional phrase required
	pp-loc-opt	locational/directional prepositional phrase optional
	pp-temp	temporal prepositional phrase required
	pp-temp-opt	temporal prepositional phrase optional
	un	unspecified

Preposition <prep>

⇒ Used to further specify syntactic frame data categories.
⇒ Value: string:

Verb particle <verbPart>

⇒ Used to further specify syntactic frame data categories.
⇒ Value: string

3.2.3.4 Semantic MONOLINGUAL Data categories

Definition <definition>

⇒ Prose definition of entry string.
⇒ Value: string

Natural Gender <natGender>

⇒ Refers to the biological gender associated with the entry string.
⇒ Value: as follows

VALUE	DESCRIPTION
m	masculine
f	feminine
un	unspecified

Semantic Type <semType>

⇒ Represents the status of the entry string with respect to a semantic type classification structure.
⇒ Value: The following values table is adapted from a proposal from Logos Corp. See Appendix II for the complete proposal.

PART OF SPEECH	VALUE	DESCRIPTION
Noun	abs	abstract, e.g., *format, rapidity, poverty, type*
	abs-ag	abstract agent, e.g., *efficiency, cause, method, goal, event*
	abs-gen	general abstract concept, e.g., *truth, idea, justice*
	abs-nonag	non-verbal abstract, e.g., *shape, condition, class, feature*
	abs-nonag-orig	non-verbal abstract origin, e.g., *reserve, lineage, origin*
	anim	animate, e.g., *manager, committee, subscriber, buyer*
	anim-ani	animal, e.g., *deer, bacteria, gnat, weasel*
	anim-hum	human, e.g., *employee, scientist, Professor, Mrs.*
	anim-hum-func	office, title, e.g., *Dr., President, General*
	anim-hum-pn	human proper name, e.g., *John, Mr. Smith, Marie*
	anim-soc	social institution, e.g., *agency, company, bureau, business*
	anim-soc-org	specific organization, e.g., *EC, United Nations, NASA*
	asp	aspective, e.g., *prototype, majority, piece*
	cnc	concrete, e.g., *table, battery, ligament, missile*
	cnc-ag	concrete agent, e.g., *camera, radio, truck, explosives*
	cnc-amor	amorphous, e.g., *breeze, tide, atmosphere*
	cnc-atom	atomistic, e.g., *electron, granule, nucleus*

PART OF SPEECH	VALUE	DESCRIPTION
	cnc-class	classifier, e.g., *compound, substance, element*
	cnc-color	color, e.g., *olive, orange, cherry*
	cnc-ednm	edible (non-mass), e.g., *cracker, lemon, pork chop*
	cnc-func	functional, e.g., *box, wall, pipe, circuit, shirt*
	cnc-light	impulse/light, e.g., *beacon, ray, tone, flare*
	cnc-mark	mark/blemish, e.g., *boil, blemish, scratch*
	cnc-nat	natural, e.g., *cloud, pebble, flower*
	cnc-nat-plant	plant, e.g., *violet, clove, lilac*
	inform	information, e.g., *newspaper, symbol, rule, ballistics*
	inform-sen	semiotic system, e.g., *address, signal, code, number*
	loc	locative, e.g., *office, zone, city, room, Munich*
	mass	mass, e.g., *iron, water, sand, fiber, fire, heat*
	mass-mat	material, e.g., *aluminum, wool, plastic, glass*
	meas	measure, e.g., *pressure, quantity, gram, rpm, voltage*
	meas-abs	abstract measure, e.g., *temperature, length, velocity*
	meas-disc	discrete measurable concept, e.g., *increment, sum, count*
	meas-unit	unit of measure, e.g., *inch, cm, hour, volt, hertz, kph*
	proc	process, e.g., *correction, analysis, call, removal*
	tmp	temporal, e.g., *summer, morning, September, Friday*
Verb	achiev	achievement
	act	unspecified activity
	emot	emotion
	event	event
	ment-act	mental activity
	mov	movement
	mov_motdir	directed motion, e.g., *dance, depart, fly, go*
	mov_motnd	non-directed motion, e.g., *depart, go, walk*
	noise	noise-producing
	phys-act	physical activity, e.g., *persist, refrain, appear*
	percept	perceptive
	perm	permission verb
	pha	phasal verb
	pro	process
	sense	sense
	situat	situation
	stat	stative, e.g., *grow, become, sound*
Adjective	color	color, e.g., *red, yellow*
	cnt	countable
	deg	degree, e.g., *acute, intense, substantial*
	indef	indefinite
	loc	locative, e.g., *above, forward, regional*
	man	manner, e.g., *charismatic, intrepid, personable*
	mea	measure, e.g., *approximate, huge, minimal*
	seq	sequence, e.g., *consecutive, daily, former*
	shape	shape
Adverb	conn	connective
	deg	degree, e.g., *merely, approximately, completely*
	freq	frequency, e.g., *again, once, twice*
	man	manner, e.g., *by hand, electronically, simultaneously*
	prob	probability, e.g., *conceivably, by chance, maybe*
	seq	sequence, e.g., *primarily, lastly, first*

PART OF SPEECH	VALUE	DESCRIPTION
	spa	space, e.g., *anywhere, to the right, inside*
	stat	stative, e.g., *alike, at ease, out of commission*
	tmp	time, e.g., *still, yet, already, at one time*
Prep	cau	causal, e.g., *as a result of, because of*
	cau-neg	causal-negation, e.g., *despite, in the absence of*
	comb	combinatorial, e.g., *with, in combination with*
	con	connective
	concess	concessive
	cond	conditional
	cor	correlative
	cor-neg	correlative-negation
	dir	direction
	incl	inclusive, e.g., *in addition to, inclusive of*
	incl-neg	inclusive-negation, e.g., *except for, instead of, without*
	instr	instrumental, e.g., *by, by means of, by way of*
	loc	locative
	loc-ext	locative-extensive
	loc-from	locative-from, e.g., *from, off of, out of*
	loc-path	locative-path
	loc-to	locative-to, e.g., *to*
	man	manner
	mea	measure
	mod	modal
	orig	origin
	path	path
	purp	purpose, e.g., *for, for the benefit of*
	qual	qualitative
	quant	quantitative
	tmp	time, e.g., *at the beginning of, during, prior to*
	tmp_ext	temporal_extensive
	tmp_from	temporal_from
	tmp_id	temporal_identical
	tmp_to	temporal_to
	unit	unit
	un	unspecified

3.2.4 Values for CROSS-REFERENCE Data categories

Cross-Reference Link Type <crLinkType>

⇒ Indicates the type of cross-reference link that pertains between the entry from which the link originates and the entry to which the link points.
⇒ Value: as follows

Cross-reference relations have been augmented by ISO relations (most of which formally apply to concepts rather than the terms themselves, but have adapted them here for the purposes of OLIF2) and the analysis contained in EuroWordNet (July, 2000).

VALUE	DESCRIPTION
synonym	synonym of
near-synonym	near synonym of
antonym	antonym of
near-antonym	near antonym of
has-hyperonym	is kind of (subordinate)
has-hyponym	has kind (superordinate)
has-holonym	part of
has-meronym	whole of
has-holo-member	member of (member-set)
has-mero-member	set (member-set)
has-holo-portion	portion of
has-mero-portion	has portion
has-holo-madeof	ingredient of
has-mero-madeof	has ingredient
has-holo-location	more specific place
has-mero-location	wider place
causes	cause of
is-caused-by	effect of
has-subevent	(between verbs/gerunds) *e.g.,* ***sleep*** *~ snore*
is-subevent-of	(between verbs/gerunds) *e.g.,* ***snore*** *~ sleep*
role	activity that something (noun) is involved in
involved	thing (noun) involved in activity represented by verb
role-agent	typical activity of agent, *e.g.,* ***teaching*** *~ teacher*
involved-agent	typical agent of activity, *e.g.,* ***teacher*** *~ teaching*
role-patient	activity undergone by patient, *e.g.,* ***learning*** *~ learner*
involved-patient	typically undergoes activity, *e.g.,* ***learner*** *~ learning*
role-result	activity that results in object, *e.g.,* ***crystallising*** *~ crystal*
involved-result	object resulting from activity, *e.g.* ***crystal*** *~ cystallising*
role-instrument	activity instrument is used for, *e.g.,* ***hammering*** *~ hammer*
involved-instrument	instrument used for activity, *e.g.,* ***hammer*** *~ hammering*
role-location	activity typical of a place, *e.g.,* ***teaching*** *~ school*
involved-location	place where activity occurs, *e.g.,* ***school*** *~ teaching*
role-direction	activity from/to/over/across/thru a place, *e.g.,* ***crossing*** *~ river*
involved-direction	place from/to/over/thru,etc. which activity occurs, *e.g.,* ***river*** *~ cross*
produces	producer of
is-product-of	product of
process-step	step in a process
in-sequence	element in a sequence
is-spatial-rel	related spatially

is-associated	associated term
is-child-of	offspring of
is-parent-of	parent of
is-used-for	is used for
use	use to which something is put
in-manner	*(verb ~ adv)* ***snore*** *~ noisily*
manner-of	*(adv ~ verb)* ***noisily*** *~ snore*
be-in-state	*(noun ~ adj)* ***tycoon*** *~ wealthy*
state-of	*(adj ~ noun)* ***wealthy*** *~ tycoon*
previous	previous version of entry
no-synonym	not allowed as synonym
has-no-syn	has disallowed synonym
is-derived-from	derivational morphology
has-derived	derivational morphology
pertains-to	*(adj ~ noun)* ***chemical*** *~ chemistry*
is-pertained-to	*(noun ~ adj)* ***chemistry*** *~ chemical*
has-instance	class
belongs-to-class	instance of class
keyword	keyword
acronym	acronym
has-acronym	has acronym
orth-variant	orthographical variant -> **see attribute table that follows**
has-orth-variant	has orthographical variant
abbreviation	abbreviated form
has-abbrev	has abbreviated form
headword	head word of compound/phrase
has-headword	has head word
fuzzynym	*(noun ~ noun; verb ~ verb)* fuzzy semantic relation
repl-controlled	replace with controlled language
Compound noun codes:	*Indicate relations between compnd nouns and compnd elements*
co-role	general relation between compound noun and compound element
co-agent-patient	*criminal ~ crime victim*
co-patient-agent	*crime victim ~ criminal*
co-agent-instrument	*guitar player ~ guitar*
co-instrument-agent	*guitar ~ guitar player*
co-agent-result	*novel writer ~ novel*
co-result-agent	*novel ~ novel writer*
co-patient-instrument	*ice ~ ice saw*
co-instrument-patient	*ice saw ~ ice*
co-patient-result	*pastry dough ~ pastry*
co-result-patient	*pastry ~ pastry dough*
co-instrument-result	*movie camera ~ movie*
co-result-instrument	*movie ~ movie camera*
un	relation unspecified

Orthographic Variant Type <orthVariantType>

⇒ Information about the type of orthographic variant that the target of a cross-reference represents.

⇒ Value: Linguatec has requested the following values to coordinate with the cross-reference link *orth-variant – has orth-variant* for German; this data category can be expanded or changed based on user requirements.

Attribute	**Description**	**Example**
german-1	Match vowels to stem	Schänke/Schenke
german-2	"selbstständig" instead of "selbständig"	unselbstständig/unselbständig
german-3	German spelling of non-German words	Soße/Sauce
german-4	Write "f" instead of "ph"	Fantasie/Phantasie
german-5	Write "r" instead of "rh"	Katarr/Katarrh
german-6	Write "t" instead of "th"	Tunfisch/Thunfisch
german-7	Write "zi" instead of "ti"	differenziell/differentiell
german-8	Plural "ices" instead of "izes"	Indices/Indizes
german-9	New spelling of non-German words	Campagne/Kampagne
german-10	Repeat three letters without a hyphen	Schifffahrt/Schiff-Fahrt
german-11	Write preposition and “weak” noun as two words	im Stande/imstande
german-12	Write “nicht” in compound adjectives as a separate word	nicht öffentlich/nichtöffentlich
german-13	Write “rein” in compound adjectives as a separate word	rein seiden/reinseiden
german-14	Write “wohl” in compound adjectives as a separate word	wohl tuend/wohltuend
german-15	Write non-German words with multiple parts as a single word	Bluejeans/Blue Jeans
german-16	Write non-German words with multiple parts with a hyphen	Fall-out/Fallout
un	unspecified	

3.2.5 Values for TRANSFER Data categories

Degree of Equivalence <equival>

⇒ The degree of transfer relationship between words/phrases in two different languages.
⇒ Value: as follows:

VALUE	DESCRIPTION
full	full equivalence
partial	partial equivalence
alt	alternate transfer
none	no equivalence
un	unspecified

For a more detailed explanation of the following data categories, see Appendix II, 'Transfer Restrictions and Structural Changes to Transfer.'

Transfer Restriction Statement <trRestrictStmt>

⇒ Container for grouping multiple, related transfer restrictions.
⇒ Value: element(s) (used as grouping construct)

Transfer Restriction <trRestrict>

⇒ Expresses a transfer restriction.
⇒ Value: element(s) (used as grouping construct)

Context Statement <contextStmt>

⇒ Indicates a logical expression about the context(s) specified in the transfer restriction or structural change
⇒ Value: element(s) (used as grouping construct)

Context <context>

⇒ Indicates 1) the context for a given translation of a source word/phrase into a target word/phrase, or 2) the context for a structural change in the target language.
⇒ Value: as follows:

VALUE	DESCRIPTION
head	the entry word itself or the head of the entry string
pp	prepositional phrase
genobj	possessive phrase, eg., "of n"
adj	descriptive/predicate adjective
prep	prep in phrase in which entry noun is prep object

VALUE	DESCRIPTION
subj	subject noun
dobj	direct object noun
iobj	indirect object noun
comp	sentential complement
adv	adverb
prepobj	noun object of preposition
string	refers to phrase that must be matched word-for-word; phrase itself is specified as value of data category <testValue>

Logical Operator <logOp>

⇒ Designates a logical operator.
⇒ Value: AND, OR, NOT

Test Statement <testStmt>

⇒ Expresses a transfer restriction.
⇒ Value: element(s) (used as grouping construct)

Test Type <testType>

⇒ Indicates whether the test on the context is of type *string* or *data category*.
⇒ Value: *STRING, DATACAT*

Test Data Category <testDC>

⇒ Names the data category to which a test pertains.
⇒ Value: valid name of OLIF data category.

Test Value <testValue>

⇒ Describes the value of the string or data category being tested on the context.
⇒ Value: string

Structural Change Statement <structChangeStmt>

⇒ Container for grouping multiple, related structural changes.
⇒ Value: element(s) (used as grouping construct)

Structural Change <structChange>

⇒ Describes a change in the target language vis-à-vis the source structure based on the transfer restriction having been satisfied.
⇒ Value: element(s) (used as grouping construct)

Change Type <changeType>

⇒ Indicates the type of change designated by the structural change
⇒ Value: as follows:

VALUE	DESCRIPTION
add-in-target	add an element in the target
del-in-target	delete an element in the target
change-vbform	change the verb form
change-role	change the role of an argument
assign-case	assign case to a noun
change-el-transfer	change the transfer of a context element

Change Part of Speech <changePOS>

⇒ Names the part of speech of an element being added or deleted.
⇒ Value: valid names for part of speech in OLIF.

Change Value <changeValue>

⇒ Describes the value of the string or data category being changed.
⇒ Value: as follows:

For additions/deletions: value is string of element being added/deleted

For changes to verb form:

VALUE	DESCRIPTION
active	target is active voice
passive	target is passive voice
causative	target is causative
reflexive	target is reflexive

For changes to role:

VALUE	DESCRIPTION
subj-dobj	subject is target direct object
dobj-subj	direct object is target subject
dobj-iobj	direct object is target indirect object
iobj-dobj	indirect object is target direct object
subj-iobj	subject is target indirect object
iobj-subj	indirect object is target subject

For changes to context element transfer: **Value is *string***

For case assignment:

VALUE	DESCRIPTION
n	nominative
g	genitive
d	dative
a	accusative
obj	objective
subj	subjective
loc	locative
prp	prepositional
inst	instrumental

Appendix I:

Proposed OLIF Handling for Transfer Restrictions and Structural Changes to Transfers

1. Transfer Restriction (trRestrict)

✓ A transfer restriction specifies a condition in the source language under which a given translation is valid.

✓ Transfer restrictions are definable for the following parts-of-speech:
- Noun
- Verb
- Adjective
- Adverb
- Preposition

✓ There are two basic components to a transfer restriction:

a) The *context(s)* for a given translation of a source word/phrase into a target word/phrase.

b) *Test(s)* on the data categories/values associated with the context

✓ The context may be:

a) The source word/phrase itself

b) Distinct context elements that occur with the source word/phrase within the clause. (These elements usu. fall within the syntactic frame defined for that particular word/phrase.) The context elements are generally categorised based on their part-of-speech.

c) Phrases that must be matched word-for-word for the condition to be satisfied, e.g., *trip the light fantastic, be in hot water.*

(Tests on context types (a) and (b) can be tests on data category values that are assigned in the lexicon, as well as data category values that are assigned in a system analysis process.)

✓ Context elements differ depending on the part-of-speech of the word/phrase:

Context elements for nouns:
- Attached prep phrase(s) = **N** PP…
- Attached possessive phrase = **N** (of) N
- Descriptive adjective = Adj **N**
- Prep in phrase in which noun is object of prep = Prep **N**

Context Elements for Verbs:

- Noun arguments = **V** N(Subj), N(DO), N(IO)
- Attached prep phrase(s) = **V** PP…
- Adverb = **V** Adv
- Predicate adjective = **V** Adj
- Sentential complement = **V** Comp

Context Elements for Adjectives:

- Head noun = **Adj** N
- Adverb = Adv **Adj**
- Attached prep phrase(s) = **Adj** PP… (predicate adjective)

Context Elements for Adverbs:

- Prep phrase = **Adv** PP

Context Elements for Prepositions:

- Noun object of prep = **Prep** N
- Prep phrase = **Prep** N PP

✓ When the context is the source word/phrase itself and the source string is a phrase, the context is referred to as the *head* of the phrase.

✓ Tests on context types (a) and (b) are tests on values for official OLIF data categories, including the following:

- Canonical form (canForm)
- Part of Speech (ptOfSpeech)
- Semantic type (semType)
- Syntactic type (synType)
- Grammatical gender (gender)
- Natural gender (natGender)
- Case (case)
- Number (number)
- Degree (degree)
- Voice (voice)
- Mood (mood)
- Tense (tense)
- Aspect (aspect)
- Subject field (subjField
- Product (product)
- Company (company)

✓ The test for the source phrase that must be matched word-for-word (context type (c)) is the context string itself.

2. The Representation of Transfer Restrictions in OLIF:

✓ A transfer restriction is represented as a statement within the transfer block of an entry.

The *transfer restriction statement* must contain one or more *transfer restrictions*, each containing a *context statement* and a *test statement.* The context statement groups one or more *contexts;* the test statement groups one or more *tests*. A test is represented as a *test type*, which specifies either a data category test or a string test, and a *test value,* which specifies the actual data category/value pair or string. If the test type is *data category*, the *test data category* is explicitly represented in the test block of the test statement.

1) For a noun entry:

```
<trRestrictStmt>
    <trRestrict>
        <contextStmt>
            <context>genobj</context>
        </contextStmt>
        <testStmt>
            <test>
                <testType>DATACAT</testType>
                <testDC>semType</testDC>
                <testValue>anim-hum</testValue>
            </test>
        </testStmt>
    </trRestrict>
</trRestrictStmt>
```

"*The transfer is valid if the possessive object of the entry noun is of semantic type animate-human"*

✓ The user may specify multiple contexts and multiple transfer tests within a single transfer restriction by using a logical operator *logOp* to represent *AND, OR,* and *NOT* relationships. In (2), for example, the test applies to both of the contexts that precede it in the context statement:

2) For a verb entry:

```
<trRestrictStmt>
    <trRestrict>
        <contextStmt>
            <context>subj</context>
                <logOp>OR</logOp>
            <context>dobj</context>
        </contextStmt>
        <testStmt>
            <test>
```

```
                    <testType>DATACAT</testType>
                    <testDC>semType</testDC>
                    <testValue>anim-hum</testValue>
                </test>
            </testStmt>
        </trRestrict>
    </trRestrictStmt>
```

"*The transfer is valid if the subject or direct object of the entry verb is of semantic type animate-human.*"

In (3), on the other hand, several transfer restrictions may be specified within a single transfer restriction statement to indicate that separate tests apply to the individual context statements that precede them:

3) For a verb entry:

```
<trRestrictStmt>
    <trRestrict>
        <contextStmt>
            <context>subj</context>
        </contextStmt>
        <testStmt>
            <test>
                <testType>DATACAT</testType>
                <testDC>number</testDC>
                <testValue>sg</testValue>
            </test>
        </testStmt>
    </trRestrict>
    <logOp>AND</logOp>
    <trRestrict>
        <contextStmt>
            <context>head</context>
        </contextStmt>
        <testStmt>
            <test>
                <testType>DATACAT</testType>
                <testDC>mood</testDC>
                <testValue>subj</testValue>
            </test>
        </testStmt>
    </trRestrict>
</trRestrictStmt>
```

"The transfer is valid if the subject of the entry verb is in the singular and the entry verb is in the subjunctive."

✓ Suggested values for the data category ***context*** are:

	VALUE	DESCRIPTION
context type (a):	head	the entry word itself or the head of the entry string
context type (b):	pp	prepositional phrase
	genobj	possessive phrase, eg., "of n"
	adj	descriptive/predicate adjective
	prep	prep in phrase in which entry noun is prep object
	subj	subject noun
	dobj	direct object noun
	iobj	indirect object noun
	comp	sentential complement
	adv	adverb
	prepobj	noun object of preposition
context type (c)	string	refers to phrase that must be matched word-for-word; phrase itself is specified as value of data category <testValue>

✓ Values for *testType* are: *DATACAT, STRING*

✓ Values for *testDC* are any valid OLIF data category names

✓ Values for *testValue* are:

⇒ If the test type is *DATACAT*, the value for *testValue* is the value of the data category specified in *testDC.*

⇒ If the test type is *STRING*, the value for *testValue* is the string being tested.

3. Structural Changes (structChange) in the Transfer

✓ Structural changes specify changes in the target translation based on a transfer restriction having been satisfied.

✓ Structural changes are definable for the following parts-of-speech:

- Noun
- Verb
- Adjective
- Preposition

✓ Structural changes often reflect what translators view as the 'addition' or 'deletion' of elements in the target (underlying this is the assumption that the translation grammar systematically specifies its 'standard' translation of

a source string which can be reordered based on lexical considerations); some structural changes reassign roles or specify a change in the value of a data category:

✓ **Typology of Structural Changes:**

Noun:

• Add preposition to context noun	=	**N** N -> **N** Prep N
• Delete preposition from attached PP; assign case/role to N	=	**N** Prep N -> **N** N
• Add determiner to N	=	**N** -> Det **N** **N** N -> **N** Det N **N** Prep N -> **N** Prep Det N
• Delete determiner from N	=	Det **N** -> **N** **N** Det N -> **N**N **N** Prep Det N -> **N** Prep N
• Add descriptive adjective	=	**N** -> Adj **N**
• Delete descriptive adjective	=	Adj **N** -> **N**

Verb:

• Add noun argument; Assign case/role to N	=	**V** -> **V** N
• Delete noun argument	=	**V** N -> **V**
• Add preposition to object N	=	**V** N -> **V** Prep N
• Delete preposition from attached PP; assign case/role to N	=	**V** Prep N -> **V** N
• Reorder cases/roles of argument N's	=	**V** N1 N2 -> **V** N2 N1
• Change voice of verb; adjust cases/roles of noun arguments	=	**V**(active) -> **V**(passive) **V**(passive) -> **V**(active)
• Add adverb	=	**V** -> **V** Adv
• Delete adverb	=	**V** Adv -> **V**
• Add predicate adjective	=	**V** -> **V** Adj
• Delete predicate adjective	=	**V** Adj -> **V**

Adjective:

• Add adverb	=	**Adj** -> Adv **Adj**
• Delete adverb	=	Adv **Adj** -> **Adj**

Preposition:

• Add determiner for noun object	=	**Prep** N -> **Prep** Det N
• Delete determiner for noun object	=	**Prep** Det N -> **Prep** N
• Add descriptive adjective	=	**Prep** N -> **Prep** Adj N
• Delete descriptive adjective	=	**Prep** Adj N -> **Prep** N

4. The Representation of Structural Changes in OLIF:

✓ Based on the typology above, there are six basic structural changes proposed:

- add element(s) in target (*add-in-target)*
- delete element(s) in target (*del-in-target)*
- change verb form *(change-vbform)*
- change argument roles *(change-role)*
- change transfer of context element *(change-el-trans)*
- assign case *(assign-case)*

✓ The *add* and *delete* structural changes require a specification of the part of speech of the element(s) being added/deleted in the target.

✓ Structural changes are grouped within *structChangeStmt* tags within the transfer block of an entry and follow any transfer restrictions that apply to them.

✓ A structural change itself is expressed as a *context statement,* consisting of one or more target *context* specifications, and a change, consisting of a *change type,* the *part of speech of an element being added or deleted,* and a *value for the change*:

4) For a noun entry:

```
<trRestrictStmt>
    <trRestrict>
        <contextStmt>
            <context>genobj</context>
        </contextStmt>
        <testStmt>
            <test>
                <testType>DATACAT</testType>
                <testDC>semType</testDC>
                <testValue>anim-hum</testValue>
            </test>
        </testStmt>
    </trRestrict>
</trRestrictStmt>
<structChangeStmt>
    <structChange>
        <contextStmt>
            <context>genobj</context>
        </context Stmt>
        <changeType>add-in-target</changeType>
        <changePOS>prep</changePOS>
        <changeValue >of</changeValue>
    </structChange>
</structChangeStmt>…….
```

"If the possessive object of the entry noun is of semantic type animate-human, the transfer is valid and the possessive object in the target should be expressed as a prepositional phrase with the preposition 'of'."

- ✓ A structural change may specify a general addition or deletion in the target, e.g., deleting the determiner in a noun phrase:

5) For a preposition entry:

```
<trRestrictStmt>
    <trRestrict>
        <contextStmt>
                <context>prepobj</context>
        </context Stmt>
        <testStmt>
                <test>
                    <testType> DATACAT</testType>
                    <testDC>synType</testDC>
                    <testValue>prop</testValue>
                </test>
        </testStmt>
    </trRestrict>
</trRestrictStmt>
<structChangeStmt>
    <structChange>
        <contextStmt>
                <context>prepobj</context>
        </contextStmt>
        <changeType>del-in-target</changeType>
        <changePOS>det</changePOS>
   </structChange>
</structChangeStmt>…….
```

"If the object of the preposition is of syntactic type proper noun, the transfer is valid and the target object of the preposition should be expressed without a determiner.'

- ✓ Multiple structural changes may be represented using the logical operator *logOp*. Unlike with transfer restrictions, only the operator *AND* is valid for a structural change:

6) For a verb entry:

```
<trRestrictStmt>
    <trRestrict>
        <contextStmt>
```

```
                <context>subj</context>
            </contextStmt>
            <testStmt>
                <test>
                    <testType> DATACAT</testType>
                    <testDC>semType</testDC>
                    <testValue>anim-hum</testValue>
                </test>
            </testStmt>
        </trRestrict>
    </trRestrictStmt>
    <structChangeStmt>
        <structChange>
            <contextStmt>
                    <context>subj</context>
            </contextStmt>
            <changeType>change-role</changeType>
            <changeValue>subj-dobj</changeValue>
        </structChange>
        <logOp>AND</logOp>
        <structChange>
            <contextStmt>
                    <context>dobj</context>
            </contextStmt>
            <changeType>change-role</changeType>
            <changeValue>dobj-subj</changeValue>
        </structChange>
    </structChangeStmt>…….
```

"If the subject of the source verb is of semantic type animate-human, the transfer is valid and the subject of the target verb is expressed as the direct object, the direct object of the target verb is expressed as the subject.""

- ✓ **Suggested values for data categories associated with structural changes:**

- ✓ For *add* and *delete*, the value for the change is the string in the target to be added/deleted.

- ✓ Values for the other changes are as follows:

For changes to verb form:

VALUE	DESCRIPTION

VALUE	DESCRIPTION
active	target is active voice
passive	target is passive voice
causative	target is causative
reflexive	target is reflexive

For changes to role:

VALUE	DESCRIPTION
subj-dobj	subject is target direct object
dobj-subj	direct object is target subject
dobj-iobj	direct object is target indirect object
iobj-dobj	indirect object is target direct object
subj-iobj	subject is target indirect object
iobj-subj	indirect object is target subject

For changes to context element transfer: **Value is *string***

For case assignment:

VALUE	DESCRIPTION
n	nominative
g	genitive
d	dative
a	accusative
obj	objective
subj	subjective
loc	locative
prp	prepositional
inst	instrumental