

The Structure and Content of the Body of an OLIF v.2 File

OLIF2 Consortium
July 3, 2001

Susan McCormick, Consultant to SAP AG
With Input from OLIF2 Consortium Members

CONTENTS

| | | |
|---------|----------------------------------------------------------------------------|----|
| 1 | General..... | 2 |
| 2 | The Structure of the Body of an OLIF v.2 File..... | 2 |
| 3 | The Content of OLIF v.2 Entries..... | 3 |
| 3.1 | Table of Data Categories..... | 3 |
| 3.2 | Values..... | 8 |
| 3.2.1 | Values for KEY Data Categories..... | 8 |
| 3.2.2 | Values for GENERAL Data Categories..... | 10 |
| 3.2.3 | Values for Optional MONO Data Categories..... | 11 |
| 3.2.3.1 | Administrative MONO Data Categories..... | 11 |
| 3.2.3.2 | Morphological MONO Data Categories..... | 14 |
| 3.2.3.3 | Syntactic MONO Data Categories..... | 18 |
| 3.2.3.4 | Semantic MONO Data Categories..... | 22 |
| 3.2.4 | Values for CROSS-REFERENCE Data Categories..... | 25 |
| 3.2.5 | Values for TRANSFER Data Categories..... | 28 |
| | APPENDIX I: Semantic Type Analysis..... | 32 |
| | APPENDIX II: Transfer Restrictions and Structural Changes to Transfer..... | 39 |

1 General

The original structure of an OLIF file, as defined for the OTELO project, was characterised by a *header*, which contained data that was relevant to all of the lexical/terminological entries in the file, and the *body*, which contained the entries themselves. We propose to maintain this basic structure with only minor changes in the second version of OLIF.

In this document, we present a description of the proposed structure and content of the body of the OLIF v.2 file. For the consortium proposal on the file header, see updated proposal documents for OLIF v.2 administrative data categories and XML representation.

2 The Structure of the Body of an OLIF v.2 File

The body of an OLIF v.2 file is a list of entries that contain data that is grouped according to the linguistic/lexical/terminological character of the information being represented. The groups are sub-lists of data category/value pairs (represented in XML as tags that reflect the element types, attributes, and values defined in the XML DTD/schema). An OLIF v.2 entry is structured as a monolingual entry with optional links to represent cross-reference and transfer relations. Accordingly, the proposed groups for an OLIF v.2 entry are:

- **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.

The OLIF v.2 entry 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 (to be published in conjunction with the SALT consortium).
- **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*.

As with the original OLIF, this set of key data categories is required in the monolingual group of the entry in order to identify the entry itself. In addition, it is used as well in any cross-reference group in the entry (with the exception of the *language* data category¹) in order to identify the entry that is pointed to in the cross-reference relation, and in any transfer group in the entry, in order to identify the entry that is pointed to in the transfer relation.

¹ For the cross-reference group, the data category *language* is not required since cross-reference relations are defined as intralingual links.

Since the specification of cross-reference and transfer links is optional, a minimal well-formed OLIF v.2 entry contains a monolingual group with the key data categories *canonical form*, *language*, *part of speech*, *subject field*, and *semantic reading*.

The proposed structure of the OLIF v.2 entry maintains the straightforwardness of the first version of OLIF, the purpose of which was to facilitate the description of a lexical/terminological entry to the extent that an NLP vendor such as Logos or Sail Labs can generate a basic, usable entry of its own from an OLIF record.

3 The Content of OLIF v.2 Entries

Data categories and values for OLIF v.2 are referred to in the tables and descriptions that follow. Data category names are, where possible, coordinated with the names of Martif data categories (ISO 12620), and generally follow Martif 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 v.2 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 technical group's proposal.)

Please note:

- Within an OLIF v.2 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.
- The current proposal specifies that the following data categories may appear ‘zero or more’ times within a group: *project*, *product*, *depSynonym*, *abbrev*, *orthvar*, *company*, *note*, *example*, *usage*

| Data category group | Data category name | Description |
|------------------------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <u>Basic:</u> <u>Obligatory</u> | | 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. |
| | <mono> | The monolingual data category delimits the monolingual data within an entry. |
| | <keyDC> | The key data category designator groups the five key data categories whose values uniquely identify an OLIF entry. |
| | <canForm> | The canonical form designates the entry string, represented in canonical form, as specified in OLIF guidelines (to be published in concert with SALT). |
| | <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 <i>canonical form</i> , <i>language</i> , <i>part of speech</i> , and <i>subject field</i> . |
| <u>General:</u> <u>Optional</u> | <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 (<i>mono</i> , <i>cross-reference</i> , or <i>transfer</i>) |
| | <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. |
| <u>Monolingual:</u> <u>Optional</u> | <monoDC> | The monolingual data category designator groups the optional data categories that may be used only within the <i>mono</i> group. |
| <i>administrative:</i> | <monoAdmin> | The monolingual administrative data category 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 <i>product name</i> , <i>trademark</i> , or <i>orthographic variant</i> . |
| | <entryFormation> | The entry formation indicates the shape/structure of the entry string. |
| | <phraseType> | Further specifies the type of phrasal entry string. |
| | <entryStatus> | Indicates the entry 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 |
| | <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. |
| <i>morphological:</i> | <monoMorph> | The monolingual morphological data category designator groups the morphological data categories within a monolingual entry. |
| | <morphStruct> | Indicates the morphological structure of the entry string. |
| | <inflection> | Encodes the inflection pattern(s) of the entry word or head 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. |
| <i>syntactic:</i> | <monoSyn> | The monolingual syntactic data category 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. |
| <i>semantic:</i> | <monoSem> | The monolingual semantic data category 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: <u>Optional</u> | <crossRefer> | The cross-reference data category defines cross-reference relations between the given entry and other entries in the lexicon in the same language. The cross-reference data category groups the cross-reference data within a monolingual entry. Within each cross-reference element, the keyDC data categories are obligatory. |
| | <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 data category holds information about the type of orthographic variant that the target of a cross-reference represents. |
| Transfer: <u>Optional</u> | <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. |
| | <equival> | The equivalence data category encodes the degree of transfer relationship between words/phrases in two different languages. |
| | <trRestrictStmt> | The transfer restriction statement data category is a container for grouping multiple related transfer restrictions. |
| | <trRestrict> | The transfer restriction data category expresses a single transfer restriction. |
| | <contextStmt> | The context statement data category is a logical expression about the context(s) specified in the transfer restriction or structural change. |
| | <context> | The context data category 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> | The logical operator data category designates a logical operator. Valid values are: AND, OR, and NOT for <i>trRestrictStmt</i> and AND for <i>structChangeStmt</i> . |
| | <testStmt> | The test statement data category states one or more tests on the context(s). |
| | <test> | The test data category states a single test. |
| | <testType> | The test type indicates the type of test. Valid values are: STRING and DATACAT. |
| | <testDC> | The data category test data category names the data category to which a test pertains. |

| | | |
|--|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | <testValue> | The test value data category 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> | A structural change describes a change in the target language vis-à-vis the source structure based on the transfer restriction having been satisfied. |
| | <changeType> | The change type data category indicates the type of change, e.g., <i>addInTarget</i> , <i>delInTarget</i> , <i>changeRole</i> , <i>assignCase</i> , etc. |
| | <changePOS> | The change part of speech data category names the part of speech of an element being added or deleted. |
| | <changeValue> | The change value data category 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 obligatorily 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.

(Again, 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 we should agree on a standard for each language, e.g., Roget's and utilize the numbering scheme for word senses from the designated standard.**

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

3.2.3.2 Morphological MONOLINGUAL Data categories

Morphological Structure <morphStruct>

- ⇒ Indicates 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

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 |
| Preposition | 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; based on formalism to be provided by Systran

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 fillers are language-specific and we don't have a formal representation of them for the languages we're covering – need a decision here on whether to handle the fillers with OLIF2 or to leave it to the vendors to construe.)

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'd together are enclosed in parentheses 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 <i>president</i> " |
| | 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 |

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

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

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

Note: In addition to the data categories specified here for the cross-reference group, users must also nest the obligatory basic data categories *canonical form*, *part of speech*, *subject field* and *semantic reading* within each set of <crossRefer> tags in the entry in order to identify the entry that is pointed to in the cross-reference relation.

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., <i>sleep</i> ~ <i>snore</i> |
| is-subevent-of | (between verbs/gerunds) e.g., <i>snore</i> ~ <i>sleep</i> |
| 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., <i>teaching</i> ~ <i>teacher</i> |
| involved-agent | typical agent of activity, e.g., <i>teacher</i> ~ <i>teaching</i> |
| role-patient | activity undergone by patient, e.g., <i>learning</i> ~ <i>learner</i> |
| involved-patient | typically undergoes activity, e.g., <i>learner</i> ~ <i>learning</i> |
| role-result | activity that results in object, e.g., <i>crystallising</i> ~ <i>crystal</i> |
| involved-result | object resulting from activity, e.g. <i>crystal</i> ~ <i>cristallising</i> |
| role-instrument | activity instrument is used for, e.g., <i>hammering</i> ~ <i>hammer</i> |
| involved-instrument | instrument used for activity, e.g., <i>hammer</i> ~ <i>hammering</i> |
| role-location | activity typical of a place, e.g., <i>teaching</i> ~ <i>school</i> |
| involved-location | place where activity occurs, e.g., <i>school</i> ~ <i>teaching</i> |
| role-direction | activity from/to/over/across/thru a place, e.g., <i>crossing</i> ~ <i>river</i> |
| involved-direction | place from/to/over/thru/etc. which activity occurs, e.g., <i>river</i> ~ <i>cross</i> |

| | |
|-----------------------|--------------------------------------------------------------------|
| 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: | <i>Indicate relations between compnd nouns and compnd elements</i> |
| co-role | general relation between compound noun and compound element |
| co-agent-patient | <i>criminal</i> ~ <i>crime victim</i> |
| co-patient-agent | <i>crime victim</i> ~ <i>criminal</i> |
| co-agent-instrument | <i>guitar player</i> ~ <i>guitar</i> |
| co-instrument-agent | <i>guitar</i> ~ <i>guitar player</i> |
| co-agent-result | <i>novel writer</i> ~ <i>novel</i> |
| co-result-agent | <i>novel</i> ~ <i>novel writer</i> |
| co-patient-instrument | <i>ice</i> ~ <i>ice saw</i> |
| co-instrument-patient | <i>ice saw</i> ~ <i>ice</i> |
| co-patient-result | <i>pastry dough</i> ~ <i>pastry</i> |
| co-result-patient | <i>pastry</i> ~ <i>pastry dough</i> |
| co-instrument-result | <i>movie camera</i> ~ <i>movie</i> |
| co-result-instrument | <i>movie</i> ~ <i>movie camera</i> |
| 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/nichtöffentliche |
| 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

Note: In addition to the data categories specified here for the transfer group, users must also nest the obligatory basic data categories *canonical form, language, part of speech, subject field and semantic reading* within each set of <transfer> tags in the entry in order to identify the entry that is pointed to in the transfer relation.

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 |
| 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 v. 2 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 v.2.

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:

Proposal for Semantic Type Data category Values Brigitte Orliac, Logos August 2000

Current OLIF semantic types: to be edited for OLIF2

1.1 Semantic type definition table

1.1.1 Nouns

I've reorganized most noun values (organized them in a tree structure). Values from the existing Semantic type definition table which have been attached to a higher level node are in italics. New values (in green) do not have a Semantic Type ID. Values in red are questionable. These values do not have a corresponding value in the SAL taxonomy. My recommendation is to get rid of them. We should also get rid of redundant values *bpart*, *cnt*, *col*, *nonh*, *prof*, and *vn* (See reason below).

| SEMANTIC TYPE ID | SHORT FORM | SEMANTIC TYPE | EXAMPLES |
|------------------|-----------------------|-------------------------------------------------------------------------------|-----------------------------------------------|
| 21 | abs | abstract | format, rapidity, poverty, type |
| 35 | <i>abs_ag</i> | <i>abstract_agent</i> | <i>efficiency, cause, method, goal, event</i> |
| | <i>abs_gen</i> | <i>abstract_general abstract concepts</i> | truth, Americanism, justice, idea |
| | <i>abs_nonag</i> | <i>abstract_non-verbal abstracts</i> | shape, condition, class, feature |
| 18 | <i>abs_nonag_orig</i> | <i>abstract_non-verbal abstracts_origin</i> | <i>reserve, lineage, origin</i> |
| 1 | age | age | |
| 23 | anim | animate | manager, committee, subscriber, buyer |
| 22 | <i>anim_ani</i> | <i>animate_animal</i> | <i>deer, bacteria, gnat, weasel</i> |
| 16 | <i>anim_hum</i> | <i>animate_human</i> | <i>employee, scientist, Professor, Mrs.</i> |
| 36 | <i>anim_hum_func</i> | <i>animate_human_office-title</i> | 1 Mr., Dr., President, General |
| 38 | <i>anim_hum_pn</i> | <i>animate_human_name</i> | 2 John, Mr. Smith, Marie |
| 31 | <i>anim_soc</i> | <i>animate_social institution</i> | <i>agency, company, bureau, business</i> |
| 40 | <i>anim_soc_org</i> | <i>animate_social institution_organisation</i> | 3 Logos, EEC, United Nations, NASA |
| | as | 4 aspective | 5 prototype, majority, piece |
| 24 | bpart | body part of living beings (delete; in concrete_agent or concrete_functional) | |
| 25 | cnc | concrete | table, battery, ligament, missile |
| 35 | <i>cnc_ag</i> | <i>concrete_agent</i> | <i>camera, radio, truck, explosives</i> |

| | | | |
|------|----------------------|-------------------------------------------------------------|------------------------------------------------|
| | cnc_amor | 6 concrete_amorphous | 7 breeze, tide, atmosphere |
| | cnc_atom | 8 concrete_atomistic | 9 electron, granule, nucleus |
| | cnc_class | 10 concrete_classifiers | 11 compound, substance, element |
| 2 | <i>cnc_clr</i> | <i>concrete_color</i> | <i>olive, orange, cherry</i> |
| | cnc_ednm | concrete_edibles (non mass) | cracker, pork chop, lemon |
| | cnc_func | 11.1.1 concrete_functional | 11.1.2 box, wall, pipe, circuit, shirt |
| | cnc_light | 12 concrete_impulses/ligths | 13 beacon, ray, tone, flare |
| | cnc_mark | 14 concrete_marks/blemishes | 15 boil, blemish, scratch |
| | cnc_nat | 16 concrete_natural | 17 cloud, pebble, flower |
| 27 | <i>cnc-nat-plant</i> | <i>concrete_natural_plant</i> | <i>violet, clove, lilac</i> |
| 3 | cnt | countable (delete; in syntactic type) | |
| 33 | col | collective (delete; implicit in animate_social institution) | |
| 15 | con | connective | |
| 12 | deg | degree | |
| 13 | dir | direction | |
| 4, 6 | ind | indefinite | |
| | 17.1 inf | 18 information | 19 newspaper, symbol, rule, ballistics |
| 30 | <i>inf_sem</i> | <i>information_semiotic_system</i> | <i>address, signal, code, number</i> |
| 5 | loc | locative | office, zone, city, room, Munich |
| 17 | man | manner | |
| | mass | 20 mass | 21 iron, water, sand, fiber, fire, heat |
| 26 | <i>mass_mat</i> | <i>mass_material</i> | <i>aluminum, wool, plastic, glass</i> |
| 6 | mea | measure | pressure, quantity, gram, rpm, voltage |
| | mea_abs | measure measurable abstract concepts | temperature, length, velocity, duration |
| | mea_dis | measure discrete measurable concepts | increment, sum, valence, count |
| 32 | <i>mea_unit</i> | <i>measure_unit</i> | <i>inch, hour, mph, volt, hertz</i> |
| 19 | 21.1 mod | modal | |
| 34 | nonh | non-human (delete; implicit in concrete, etc.) | |
| 28 | pot | potent | |
| 29 | pro | process | correction, analysis, call, removal |

| | | | |
|----|-------|---------------------------------------------|---------------------------------------|
| 37 | prof | job title (delete; see anim_hum_func above) | |
| 11 | qual | qualitative | |
| 14 | quant | quantitative | |
| 7 | seq | sequence | |
| 8 | shp | shape | |
| 9 | siz | size | |
| 20 | spa | spatial | |
| 10 | tmp | temporal | summer, morning, September, Friday |
| 39 | vn | first name (delete; see anim_hum_pn above) | |

1.1.2 Verbs

Intransitive verbs are adequately described with values *mov*, *motdir*, *motnd*, *pact* and *stat*. I've reorganized *movdir* and *motnd* (attach them to *mov*). Transitive verb SALs are strongly syntactic (supported by Syntactic frame). SALs which fall outside of Syntactic frame are too idiosyncratic for values in existing Semantic type: a distinction made between two types of transitive verbs, the reciprocal (a) and the objective (b), is based on the perceived effect of the verb on its arguments: effect on both subject and object in (a), on object alone in (b). Values in red below do not have a corresponding value in the SAL taxonomy. I really don't know how to organize them in a tree structure. My recommendation is to leave them in (for now).

| SEMANTIC TYPE ID | SHORT FORM | SEMANTIC TYPE | EXAMPLES |
|------------------|-------------------|-------------------------------------|---------------------------------|
| 70 | achiev | achievement | |
| 71 | act | unspecified activity | |
| 68 | emot | emotion | |
| 75 | event | event | |
| 73 | mact | mental activity | |
| 65 | mov | movement | dance, depart, fly, go |
| 63 | <i>mov_motdir</i> | <i>movement_directed motion</i> | <i>depart, go, walk</i> |
| 64 | <i>mov_motnd</i> | <i>movement_non-directed motion</i> | <i>dance, sail</i> |
| 67 | noise | noise producing | |
| 72 | pact | physical activity | <i>persist, refrain, appear</i> |
| 66 | percept | perceptive | |
| 62 | perm | permission verb | |
| 61 | pha | phasal verb | |
| 29 | pro | process | |
| 69 | sen | sense | |
| 76 | situat | situation | |
| 74 | stat | stative | <i>grow, become, sound</i> |

1.1.3 Adjectives

Adjective SALs are strongly syntactic (supported by Syntactic frame). Adjectives in the SAL high level node *Adverbial* (always have an adverb counterpart) are adequately described with values *deg*, *loc*, *man*, *mea*, and *seq* (I applied values *deg* and *man* to adjectives also).

Adjectives in other SAL nodes are supported by Syntactic position, Syntactic type, and Syntactic frame. Participial adjectives were not represented. I added values for them in Syntactic type. Values in red are questionable. These values do not have a corresponding value in the SAL taxonomy. My recommendation is to get rid of them. We should also get rid of redundant value *siz* (See reason below).

| SEMANTIC TYPE ID | SHORT FORM | SEMANTIC TYPE | EXAMPLES |
|------------------|------------|---------------|------------------------------------|
| 1 | age | age | |
| 2 | clr | color | |
| 3 | cnt | countable | |
| 12 | <i>deg</i> | <i>degree</i> | <i>acute, intense, substantial</i> |
| 16 | hum | human | |
| 4, 6 | ind | indefinite | |

| | | | |
|----|-----|---------------------------|------------------------------------------|
| 5 | loc | locative | above, forward, regional |
| 17 | man | manner | <i>charismatic, intrepid, personable</i> |
| 6 | mea | measure | approximate, huge, minimal |
| 7 | seq | sequence | consecutive, daily, former |
| 8 | shp | shape | |
| 9 | siz | size (delete; in measure) | |

1.1.4 Adverbs

Adverbs in the SAL high level node *Locative* are adequately described with values *freq*, *seq*, *spa*, *stat*, and *tmp* (I added new value *freq* and applied values *spa* and *stat* to adverbs also). Adverbs in the SAL high level node *Non-locative* are also adequately described with values *deg*, *man*, and *prob* (I added new value *prob*). Adverbs in other SAL nodes are supported by Syntactic position and/or Syntactic type. Values in red are questionable. These values do not have a corresponding value in the SAL taxonomy. My recommendation is to get rid of them. We should also get rid of redundant value *dir* (See reason below).

| SEMANTIC TYPE ID | SHORT FORM | SEMANTIC TYPE | EXAMPLES |
|------------------|------------|------------------------------|------------------------------------------|
| 15 | con | connective | |
| 12 | deg | degree | merely, approximately, completely |
| 13 | dir | direction (delete; in space) | |
| | freq | frequency | again, once, twice |
| 17 | man | manner | by hand, electronically, simultaneously |
| 19 | mod | modal | |
| 18 | orig | origin | |
| | prob | probability | conceivably, by chance, certainly, maybe |
| 11 | qual | qualitative | |
| 14 | quant | quantitative | |
| 7 | seq | sequence | primarily, lastly, first |
| 20 | spa | space | <i>anywhere, to the right, inside</i> |
| 74 | stat | stative | <i>alike, at ease, out of commission</i> |
| 10 | tmp | time ² | still, yet, already, at one time |

1.1.5 Prepositions

I've reorganized *loc_from* and *loc_to* (attach them to *loc*). I don't know what to do with the values in red (the other values have corresponding values in the SAL taxonomy). My recommendation is to leave them in (for now). Values below only represent a subset of the SAL values for prepositions. As with nouns and verbs, I didn't think I could "dump" the complete set of SAL values for this part of speech, mostly because of proprietary issues, but also because the extra SAL values are very often specific to Logos (non standard values). The result is a compromise which I hope we can revisit (and update) in the future.

| SEMANTIC TYPE ID | SHORT FORM | SEMANTIC TYPE | EXAMPLES |
|------------------|------------|----------------------|----------------------------|
| 41 | cau | causal | as a result of, because of |
| 42 | cau-neg | causal-negation | despite, in the absence of |
| 43 | comb | combinatorial | with, in combination with |
| 15 | con | connective | ??? |
| 50 | concess | concessive | ??? |
| 44 | cond | conditional | ??? |
| 45 | cor | correlative | ??? |
| 46 | cor-neg | correlative-negation | ??? |
| 13 | dir | direction | ??? |

² Links to Aspect.

| | | | |
|----|-----------------|----------------------|---------------------------------------|
| 47 | incl | inclusive | in addition to, inclusive of |
| 48 | incl-neg | inclusive-negation | except for, instead of, without |
| 49 | instr | instrumental | by, by means of, by way of |
| 5 | loc | locative | |
| 51 | loc_ext | locative_extensive | ??? |
| 52 | <i>loc_from</i> | <i>locative_from</i> | <i>from, off of, out of</i> |
| 53 | loc_path | locative_path | ??? |
| 54 | <i>loc_to</i> | <i>locative_to</i> | <i>to</i> |
| 17 | man | manner | ??? |
| 6 | mea | measure | ??? |
| 19 | mod | modal | ??? |
| 18 | orig | origin | ??? |
| 55 | path | path | ??? |
| 56 | purp | purpose | for, for the benefit of |
| 11 | qual | qualitative | ??? |
| 14 | quant | quantitative | ??? |
| 10 | tmp | time | at the beginning of, during, prior to |
| 57 | tmp_ext | temporal_extensive | ??? |
| 58 | tmp_from | temporal_from | ??? |
| 59 | tmp_id | temporal_identical | ??? |
| 60 | tmp_to | temporal_to | ??? |
| 32 | unit | unit | ??? |

1.1.6 Conjunctions

None of the values have corresponding values in the SAL taxonomy. I recommend to leave them in (for now). Conjunction SALs are too idiosyncratic for values in existing Semantic type: for example, SAL values distinguish two classes of disjunctive conjunctions (*yet, but, than, etc.*): the disjunctive conjunctions which are also adverbs, and the disjunctive conjunctions which are not. Also, unique SALs exist for all subordinating conjunctions.

| SEMANTIC TYPE ID | SHORT FORM | SEMANTIC TYPE | EXAMPLES |
|------------------|------------|---------------|----------|
| 15 | con | connective | ??? |
| 12 | deg | degree | ??? |
| 13 | dir | direction | ??? |
| 5 | loc | locative | ??? |
| 17 | man | manner | ??? |
| 19 | mod | modal | ??? |
| 18 | orig | origin | ??? |
| 11 | qual | qualitative | ??? |
| 14 | quant | quantitative | ??? |
| 7 | seq | sequence | ??? |
| 10 | tmp | time | ??? |

1.2 Syntactic Position

I need to add the following values (in green) to further categorize degree adverbs.

| PART OF SPEECH | VALUE | DESCRIPTION | EXAMPLES |
|----------------|----------|-------------------------------|-------------------------------------|
| Adverb | cl-final | clause-final | |
| | cl-init | clause-initial | |
| | deg-post | degree adverb after morpheme | apiece, enough |
| | deg-pre | degree adverb before morpheme | less, very, hardly, slightly, fully |
| | postverb | following main verb | |
| | preverb | before main verb | |

1.3 Syntactic Type

Noun and adverb values in red are redundant (See reason below). My recommendation is to get rid of them. I also need to add the following values (in green) to further categorize adjectives.

| PART OF SPEECH | VALUE | DESCRIPTION | EXAMPLES |
|-----------------------|----------|-------------------------------------------------|------------------------------|
| Noun | cnt | countable noun | |
| 21.2 | coll | collective noun (delete; in semantic type) | |
| | mass | mass noun (moved to semantic type) | |
| | mass-cnt | countable mass noun (delete) | |
| | prop | proper noun | |
| | quant | quantitative noun | |
| 21.3 Adjective | -able | -ABLE participle | consignable, obtainable |
| | attrib | attribute adjective | |
| | past | past participle | seen, revised, displayed |
| | pred | predicative adjective | |
| | pres | present participle | hoping, designing, realizing |
| Adverb | adjmod | modifies adjective | |
| | advmod | modifies adverb | |
| | clsmod | modifies clause | |
| | detmod | modifies determiner (delete; implicit in npmod) | |
| | npmod | modifies noun phrase | |
| | numod | modifies numeral | |
| | prepmod | modifies preposition | |
| | quantmod | modifies quantifier (delete; implicit in npmod) | |

1.4 Transitivity Type

I need to apply all values in Transitivity type to nouns and verbs (to adequately categorize process nouns).

| 21.3.1.1 PART OF SPEECH | VALUE | DESCRIPTION | |
|-------------------------|---------|---------------------------|---------------------|
| Noun, Verb | ditrans | ditransitive | give, provide, send |
| | intr | intransitive | |
| | mid | middle | |
| | refl | reflexive | |
| | trans | transitive | |
| | un | unspecified | |
| | unacc | unaccusative intransitive | |
| | unerg | unergative intransitive | |

1.5 Syntactic Frame

Values *pp* and *pp-loc* should link to Preposition.

| PART OF SPEECH | 21.4 VALUE | DESCRIPTION | EXAMPLES |
|-----------------------|---------------------|------------------------------------------------------|----------|
| Noun, Verb, Adjective | pp ³ | prepositional phrase required | |
| | pp-opt | prepositional phrase optional | |
| | pp-loc ⁴ | locational/directional prepositional phrase required | |

³ Links to Preposition.

⁴ Links to Preposition.

| | | | |
|--|------------|------------------------------------------------------------|--|
| | pp-loc-opt | locational/directional prepositional phrase optional | |
|--|------------|------------------------------------------------------------|--|

Appendix II:

Proposed OLIF v.2 Handling for Transfer Restrictions and Structural Changes to Transfers

(Originally from OTELO report for WP LR 1.2)

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 v.2 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 v.2:

- ✓ 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 v.2 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; = N Prep N -> N N
assign case/role to 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 -> NN
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; = V Prep N -> V N
assign case/role to N
- Reorder cases/roles of argument N's = V N1 N2 -> V N2 N1
- Change voice of verb; = V(active) -> V(passive)
adjust cases/roles of noun arguments 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 v.2:

- ✓ 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> DATA CAT </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 |
|-----------|-------------------------|
| 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 |
| | |