The RELISH Schema:
an interchange format for lexicons

Helen Aristar-Dry
Institute for Language Information and Technology
Eastern Michigan University
RELISH Meets LOEWE, Oct 10, 2011
ISO 24613: LMF (Lexical Markup Framework)

- Exists as a data model
  - no official serialization (i.e., no ordered format for machine-processing)
  - Various examples and DTD are informative, not normative
Structure of an LMF Entry

Lexical Entry
@Part of Speech

Lemma
FormRepresentation

Word Form
@phonologicalForm
@grammaticalTense...
@grammaticalNumber...

Stem

Related Form

Sense
Definition
Statement
SenseExample
SenseRelation
LIFT

- LIFT = Lexicon Interchange Format
- XML schema for lexical information (born serial)
- Developed by SIL International
- Designed to be easy to convert into and out of MDF and Fieldworks formats
Programs that support LIFT

- **WeSay** uses LIFT as its primary format.
- **FieldWorks Language Explorer (FLEEx)** can import and export LIFT files.
- **Lexique Pro** can open LIFT documents for viewing, printing, and making web pages. It can also save to LIFT format.

Utilities for LIFT

- **Solid** can convert basic SFM (standard format markers, e.g. Toolbox format) to LIFT (see: [http://lingtranssoft.info/apps/solid](http://lingtranssoft.info/apps/solid))
- **LiftTweaker** Can selectively modify a LIFT file for targeting different audiences
Lexicons in LIFT

- LIFT chosen as upload format for LEGO because of the large number of lexicons potentially available in LIFT
  - About 50 published lexicons in Lexique Pro
  - 180+ lexicons in Fieldworks Language Explorer (FLEEx) ?
  - 300+ lexicons in Shoebox/Toolbox ?
- With the owner’s permission, these could easily be integrated into the LEGO system
Structure of a LIFT Entry

LIFT Entry

Lexical Unit

Note

Sense*

Grammatical-Info

Definition

Example

Note

Note

Variant

Relation
Notes on LIFT

- Grammatical Info is attached to Sense, not to Entry or Form (differs from LMF)
- Variant is attached to Entry, not ‘sense’ – can’t add a ‘sense’ or grammatical info to a variant
- Multiple senses and variants allowed
- Highly customizable: Field, Type, and Range can be added to virtually any element (can be defined in the document header)
LL-LIFT

- Lack of constraint on use of Field and Trait constituted a problem for LEGO
- Developed ‘LL-LIFT’
  - a constrained form of LIFT
  - which still validates against the LIFT schema
Major constraints

- Grammatical Information confined to a single element (1st occurrence of Sense)
  - Delimited within db field
  - Parsed out during GOLD mapping
- Minor entries, comparison forms, inflected forms, etc
  - Separate entries
  - Linked via ‘Relation’ element
  - Unified again in the display
From LIFT to LL-LIFT

- Constraints:
  - Restriction of user-defined elements:
    - <trait> restricted to <grammatical-info>
    - <field> not allowed
  - Restricted to only one instance of <grammatical-info> per entry (only first Sense can have <grammatical-info> tag)
  - Restricted use of <gloss> to forms that have both <definition> and <gloss>
From LIFT to LL-LIFT

- Constraints (cont.):
  - Restricted note types to a limited set
  - Restricted notes to Entry, Sense, Example

- Other
  - Allowed multiple language codes per entry—will use a mul (multiple) ISO code, which is specified in the Header as referring to specific langs
  - Added metadata (OLAC and LEGO) to Header – metadata will become part of the LIFT standard
Goal: An LL-LIFT-based serialization of LMF

- One-to-one mapping of most elements
- Addition of datcat references
- Addition of TEI feature structures
LMF Lexical Entry

@Part of Speech

Lemma
Form Representation
@geographicalVariant

Word Form
@phoneticForm . . .
@grammaticalGender
@grammaticalTense . . .

Stem

Related Form

Sense
Sense Example
Sense Relation
Definition
Statement

LIFT Entry

Lexical Unit

Note

Sense
Grammatical-Info (POS)
Definition
Example
Note
Note

Variant
Relation
Grammatical information

- Aggregated by different elements:
  - Sense in LL-LIFT
  - Lemma (POS) and Word Form (@gender, @number) in LMF
- But
  - LL-LIFT restricts <grammatical-info> to 1 instance per entry
  - LMF restricts POS to 1 instance per Lemma and other grammatical attributes to 1 set per word-form
- So one-to-one mapping is possible.
- “We can consider LL-LIFT just a weird serialization of LMF” (Simons, 2010)
The Way to Harmony (3)

- No place for notes in LMF, but:
  - LMF allows features on almost all elements
  - So we treat notes as features (or feature structures)
- Added TEI feature structures to allow sub-grouping of feature content
The Way to Harmony (4)

- Added DatCat references (ISOCat Registry)
  - Required by LMF
  - Will allow us to preserve LEGO GOLD mappings (GOLD now in ISOCat Registry)
RELISH Schema

- Still under development
- May be expanded to represent a full serialization of LMF (i.e. not restricted to LIFT elements)
- Current version available
  - from the RELISH project page on LINGUIST LIST: [http://linguistlist.org/projects/relish.cfm](http://linguistlist.org/projects/relish.cfm)
  - from the GOLD Community site: [http://linguistics-ontology.org](http://linguistics-ontology.org)
RELISH Schema

- Still under development
- May be expanded to represent a full serialization of LMF (i.e. not restricted to LIFT elements)
- Current version available
  - from the RELISH project page on LINGUIST LIST: http://linguistlist.org/projects/relish.cfm
  - from the GOLD Community site: http://linguistics-ontology.org
Result?

Harmony