

Encoding samples and questions

Samples

VERB (zwei Infinitivformen + zwei Statusformen + Stativ)¹

```
<superEntry>
  <entry n="1">
    <form ><orth>ἰκᾶν</orth></form>
    <form><orth>ἰκᾶν</orth></form>
    <gramGrp><pos>Vb.</pos></gramGrp>
    <form type="inflected">ἰκᾶν</form> <gramGrp> <pos>Vb.</pos> <subc>Status nominalis</subc> </gramGrp></form>
    <form type="inflected">ἰκᾶν</form> <gramGrp> <pos>Vb.</pos> <subc>Status pronominalis</subc> </gramGrp></form>
  <entry n="2">
    <form> ἰκᾶν </form>
    <gramGrp><pos>Vb.</pos> <subc>Qualitativ</subc></gramGrp>
</superEntry>
```

PRÄPOSITION (zwei Statusformen in einem Dialekt) +

```
  <entry n="1" type="hom">
    <form type="inflected"> <usg>BS (rare)</usg><orth>ἐ</orth></form><gramGrp><subc>Status nominalis</subc></gramGrp></form>
    <form type="inflected"> <usg>BS (rare)</usg><orth>ἐ</orth></form><gramGrp><subc>Status pronominalis</subc></gramGrp></form>
    <gramGrp><pos>Präp.</pos></gramGrp>
  .... </entry>
```

SATZKONVERTER

```
<superEntry>
```

¹ If qualitative is a separate entry, then cross-reference to the infinitive is necessary.

```

<entry>
  <form>ḐṯḐ</form>
  <gramGrp><pos>Satzkonverter</pos><subcat>Relativkonverter</subcat></gramGrp>
  <form type="inflected">Ḑṯ- <gramGrp> <pos> Satzkonverter </pos> <subc>Status nominalis</subc> </gramGrp></form>
</superEntry>

```

SELBSTÄNDIGE PERSONALPRONOMEN

```

<superEntry>
<entry n="1">
  <form> ṇṯok </form>
  <form type="inflected"> ṇṯk-<gramGrp> <pos> Selbst. Pers. Pron.</pos> <subc>Status nominalis</subc> </gramGrp></form>
  <gramGrp><pos> Selbst. Pers. Pron.</pos><subcat>2. Pers.</subcat> <number>sg.</number><gen.>m.</gen.></gramGrp>

```

```

<entry n="2">
  <form> ṇṯo </form>
  <form type="inflected"> ṇṯe-<gramGrp> <pos> Selbst. Pers. Pron.</pos> <subc>Status nominalis</subc> </gramGrp></form>
  <gramGrp><pos> Selbst. Pers. Pron.</pos><subcat>2. Pers.</subcat> <number>sg.</number><gen.>f.</gen.></gramGrp>
</superEntry>

```

SUFFIXPRONOMEN

```

<entry>
  <form>-ṯ </form>
  <gramGrp><pos> Suffixpronomen</pos><subcat>1. Pers.</subcat> <number>sg.</number></gramGrp>
  <form type="inflected">-ṯ<gramGrp> <pos> Suffixpronomen</pos> <subc> nach Konsonanten und Doppelvokalen </subc> </gramGrp></form>

```

```

<entry>

```

VERBALPRÄFIXE

```

<entry>
  <form> ṽṽṽ(ṽṽ)- </form>
  <gramGrp><pos>Verbalpräfix</pos><subcat> Aorist ṽṽṽ(ṽṽ)-</subcat></gramGrp>
  <form type="inflected"> ṽṽṽ-<gramGrp> <pos> Verbalpräfix </pos> <subc> Status pronominalis </subc> </gramGrp></form>

```

<entry>

Compounds:

<entry>	<form type="compound">ταλλο (χρ-)</form>	<pos>Vb.</pos><gram type="collocPrep">χρ-</gram>	<quote>load with</quote>
<entry>	<form>ταλλο ερραι</form>	<pos>Vb.</pos><gram type="collocAdv">ερραι</gram>	<quote>rise, raise up, add</quote>
<entry type="compound">	<form>ταλλο ερραι</form>	<pos>Subst.</pos><gram type="collocAdv">ερραι</gram>	<quote>Addition</quote>

Questions

	Frage	Antwort	Kodiert in Schema
1.	<entry type="..."> <xs:enumeration value="hom"/> <xs:enumeration value="compound"/> <xs:enumeration value="foreign"/> Dürfen die „entry types“ kombiniert werden? Wenn nicht, dann was soll hinein wenn sowohl „compound“ als auch „foreign“?		In Schema wird zur Zeit nur ein „type“ zugelassen.
2.	<entry type="hom"> Entry type="hom" steht manchmal bei der 1. Entry. Wird „hom“ im Bezug auf die erste entry von <superentry> definiert? Oder sollen alle homographen „hom“ haben?	Laut TEI sollten alle Homographen mit „hom“ markiert werden. <entry type="hom" n="1"> wäre dann das „headword“.	„homs“ sind inkonsistent markiert. Eigentlich sollten wir „homs“ loswerden
3.	<form type="..."> <xs:enumeration value="lemma"/> <xs:enumeration value="inflected"/> <xs:enumeration value="compound"/> Dürfen die Typen kombiniert werden? Wenn nicht, dann was wenn sowohl „compound“ als auch „foreign“?		Nein
4.	<form type="lemma"> <i>Inf. + Noun</i> <i>Stat. + Noun</i> <form> ρ. Δοι </form> <form type=lemma> ερε</form>	We don't need type=lemma in the latter case. But we leave it for now until the final decision.	

	<p>BUT:</p> <p>Noun + Noun</p> <p>Pref. + Verb</p> <p><form type=lemma> ηλ λχo / πεqχω</p> <p>Do we need “type=lemma” in the latter case? The whole entry is anyway a lemma.</p>		
5.	Compounds		
	<p>With compounds generally there is no reference to the head word: χo, Χoi, Χω usw. Reference takes place only through “superentry” and “entry n=1”. Reference to verbal prefix through: <form type=lemma> ειπε. Reference to nominal prefix through: <xr type="cf"> <ref target="#ηντ-">Nominalpräfix</p>	<p>Two possible solutions for unification:</p> <p>1. <entry type=compound> <form> ηλλχo</form> <form type=lemma> χo</form> <form type=lemma> ηλλ</form> </entry> OR:</p> <p>2. <entry type=compound> <form type=lemma> ηλλχo</form> <xr type="cf"> <ref target="#ηντ-">Nominalpräfix</ref></xr><xr type="cf"> <ref target="# χo">Subst.</ref></xr></entry></p> <p>I prefer the second one, but it can only be done when entries have stable IDs.</p>	
6.	Question : How to deal with <cit type="example">, which has its own <usg>, <quote>, <xr>, <etym> etc. (I temporarily commented them out)		
	<pre> <entry n="2" type="hom"> <form> <usg type="geo">S B</usg> <orth>λN</orth> <gramGrp> <pos>Nominalpräfix</pos> </gramGrp> </form> <sense> <cit type="translation" xml:lang="de"> <quote>Anführer ("Großer von . . .)</quote> </cit> <cit type="translation" xml:lang="en"> </pre>	Extend the schema?	

	<pre> <quote>great one (of . . .)</quote> </cit> <cit type="translation" xml:lang="fr"> <quote>grand, chef de . . .</quote> </cit> <cit type="example"> <usg type="geo">S B</usg> <quote>ἄνωγο</quote> <cit type="translation" xml:lang="de"> <quote>Tausendschaftsführer</quote> </cit> </cit> <xr type="cf"> <ref target="#αο">tausend</ref></xr> <etym> <gloss>Chiliarch</gloss> <ref type="Greek">χιλίαρχος</ref> <ref type="WCNde" target="#819"/> </etym> <cit> <bibl>CD 10b; CED 7-8; KoptHWb 7; DELC 12; ChLCS 4a</bibl> </cit> </sense> <etym> <ref type="WCNae" target="#37460"/> <ref type="WCNde" target="#825"/> </etym> </entry> </pre>		
7.	Aorist II ????		
	<pre> <entry> <form> <usg type="geo">A</usg> <orth>ἀγαπε-</orth> </form> <gramGrp> <pos>Verbalpräfix</pos> <subc>Status nominalis</subc> </gramGrp> </pre>	<pre> "GramGrp" should contain only information related to „Satzkonverter“; “Aorist” is referenced through “xr”’: <xr type="cf"> <ref target="#αγαπε->Verbalpräfix Aorist</ref> </xr> Information about “Aorist” is also </pre>	<pre> <entry> <form> <usg type="geo">A</usg> <orth>ἀγαπε-</orth> </form> <gramGrp> <pos>Satzkonverter</pos> <subc>Status nominalis</subc> </gramGrp> </pre>

```

</form>
<form type="inflected">
  <usg type="geo">A</usg>
  <orth n="1">аѣꙗꙑ(ѣ)</orth>
  <gramGrp>
    <pos>Verbalpräfix</pos>
    <subc>Status pronominalis</subc>
  </gramGrp>
</form>
<gramGrp>
  <pos>Verbalpräfix</pos>
  <subc>Aorist ѡꙗꙑ(ꙑѣ)</subc>
  <note>Verbalpräfix Satzkonverter + Aorist </note>
</gramGrp>
<xr type="cf"> <ref target="#ѣꙗꙑѣ">Verbalpräfix Aorist II</ref></xr>
<sense>
  <cit>
    <bibl>KoptHWb 18</bibl>
  </cit>
</sense>
</entry>

```

encoded in "def"

```

  <cit type="translation"
xml:lang="de"><def>Verbalpräfix
Umstandsatzkonverter +
Aorist</def></cit>

```

```

</form>
<form type="inflected">
  <usg type="geo">A</usg>
  <orth n="1">аѣꙗꙑ(ѣ)</orth>
  <gramGrp>
    <pos>Satzkonverter</pos>
    <subc>Status
pronominalis</subc>
  </gramGrp>
</form>
<gramGrp>
  <pos>Satzkonverter</pos>

  <subc>Umstandsatzkonverter</subc>
  </gramGrp>
  <xr type="cf">
    <ref target="#ѣꙗꙑѣ-
">Verbalpräfix Umstandsatzkonverter
+ Aorist</ref>
  </xr>
  <xr type="cf">
    <ref target="#ѡꙗꙑѣ-
">Verbalpräfix Aorist</ref>
  </xr>
  <sense>
    <cit type="translation"
xml:lang="de"><def>Verbalpräfix
Umstandsatzkonverter +
Aorist</def></cit>
    <cit type="translation"
xml:lang="en"><def></def></cit>
    <cit type="translation"
xml:lang="fr"><def></def></cit>
    <cit>
      <bibl>KoptHWb 18</bibl>
    </cit>
  </sense>

```

			</sense> </entry>
8.	Bibliographic information other than a dictionary, for example to form?		
	?		
9.	What if "sense" is dialectal – new entry?		
	<pre> <entry n="3" type="hom"> <form> <usg type="geo">A L</usg> <orth>αλ</orth> </form> <form> <usg type="geo">L</usg> <orth>αλ</orth> </form> <usg type="geo">S A</usg> <orth>αλ</orth> <gramGrp> <pos>Partikel</pos> <subc>formreduziertes perfektives Partizip, immer mit Relativkonverter verwendet</subc> </gramGrp> </form> <xr type="cf"> <ref target="#ερ-">perfektives Relativpräfix</ref></xr> <sense> <cit type="example"> <quote>πεταλινε</quote> </cit> <note>vgl. gr. ὁ εὐρών (Prov 12,2)</note> </sense> </entry> </pre>		allow „usg“ for sense also?

	<pre> </cit> <cit> <bibl>CD 24a; CED 17; KoptHWb 17, 491; DELC 22; ChLCS 6b</bibl> </cit> </sense> </entry> </pre>		
10.	Relativkonverter oder Verbalpräfix?		
	<pre> <entry> <form> <orth n="1">εΝΤ-</orth> <orth n="2">ΝΤ-</orth> </form> <gramGrp> <pos>Satzkonverter</pos> <subc>Relativkonverter</subc> </gramGrp> <xr type="cf"> <ref target="#ε-">Relativkonverter</ref></xr> <xr type="cf"> <ref target="#εΤ-">Relativkonverter</ref></xr> <xr type="cf"> <ref target="#ερ-">Relativkonverter (perfektiv)</ref></xr> <xr type="cf"> <ref target="#α-">Verbalpräfix Perfekt I</ref></xr> <sense> <cit type="translation" xml:lang="de"> <quote>Relativkonverter vor Perfekt I</quote> </cit> <cit type="translation" xml:lang="en"> <quote>relative converter of Perf. I</quote> </cit> <cit type="translation" xml:lang="fr"> <quote>particule relative devant le 1er parfait</quote> </cit> </sense> <cit> <bibl>KoptHWb 37, 125; DELC 144; ChLCS 11a</bibl> </cit> </pre>		

	</sense>		
11.	T-Kausativ?		
	<entry> <form> <usg type="geo">B</usg> <orth n="1">θογελο</orth> <orth n="2">τογλο</orth> <gramGrp> <pos>Vb.</pos> <subc>Infinitiv</subc> <note>T-Kausativ</note>		
	Problem with wncoding converters, which are 1) sentence converters 2) verbal prefixes, which are combined with temporal prefixes (nschare, see below)		
12.	τη- is actually a quantifier		
	<entry> <form>τη</form> <gramGrp> <pos>Adj.</pos> </gramGrp> <sense> <cit type="translation" xml:lang="de"> <quote>jeder (nur mit in Verbindung mit ρομπε, z. B. in τηρομπε, "alljährlich")</quote>		

For TLA:

- For different senses separate entries: entry 1 NCA (hinter), entry 2 (gegen)
- For different POS separate entries: entry 1 NKOTK (Subst), entry 2 NKOTK (Verb) (or τηβοο - make pure, purity)
- Transitive / Intransitive meaning (with the same form: τηβοο - be pure, make pure) entered as same <entry> but <sense n=1>, <sense n=2>.

To do:

Task	Comment	Final version done?
- Check bugs of Emma Manning		
- check all entries where CDO says "NONE": search for nothing	I think this is in Manning's list	
- Fill out all <cit type="translation"> with empty values: French, German	Sent to Frank	
- do not allow <bib> to be in separate <sense> - all in one sense!	Adopted Schema, so that "sense" has to have min 4 "cits"	
-		

Practical recommendations for retrieving data:

Spelling:

- look in <form><orth> first (note that there could be multiple <orth>s. If its not there, then straight in <form>. The additional information can be contained in <note> - check for it too. There can be multiple forms

Note that some forms can contain brackets: $\alpha(\epsilon)\beta\eta\lambda$. – So look for forms with and without brackets

Grammar information:

- grammatical information can refer to <form> or to the whole <entry>. Thus look for <gramGrp> in <form> first. Then combine it with <gramGrp> information in <entry>. If both contain <pos>, remove one (it may happen that <pos> is encoded twice with the same value).

Combine with <note> information from <gramGrp> (if present), because for example with "converters" or "verbal prefixes" only either information type could be encoded in <pos>, another goes into <gramGrp><note> or in <sense>: Information in <note> is in English (see for example <orth> $\eta\alpha\rho\epsilon$ </orth> - Relativkonverter Aorist

Note:

- check if each type of information you are retrieving has <note>. <form> may have it, <gramGrp> may have it etc. There is useful information in there sometimes.

Compounds:

Most compounds have attribute "type=compound" in <entry>. The "head" word in the compound is usually the first <entry> in the <superentry> to which compound entry belongs. I am not a big fan of such solution. Sometimes the "head" is referenced via <xr> tag. We will have to do it for all compounds as soon as we have stable entry IDs (currently reference is done to spelling).

Collocations verb + preposition/adverb have an attribute "type=compound" in <form>. The preposition/adverb is listed in <gramGrp> tag: <gram type="collocPrep">εχN- </gram>

Possibly for parsing:

Type="hom" is useless

delete type="hom"