

Data model for citations in dictionaries

Katrien Depuydt, Jesse de Does

Abstract

In a dictionary article, lexicographers cite other resources to substantiate their analysis. These citations can be found everywhere in the dictionary, eg. in the etymological section of a dictionary article, in the definition of an entry etc.

A special form of citations are those that function as examples to give evidence, elucidate meaning or illustrate features (spelling variation, syntax, collocation, register (etc)). These citations are a representative selection of occurrences of the headword from either a corpus of texts or a corpus of snippets from texts. These citations, i.e. attestations, usually occur in the form of a quote with bibliographical information. In some cases, only a few context words are given, not the actual attested word form, in other cases, only the bibliographical reference is given.

Attestations are attributed examples in a dictionary. There are usage examples that are not attributed. They are either invented or based on corpus material, but adapted/simplified by the lexicographer. We will focus on attestations.

1 Introduction

Lexicographers use examples to support their analysis of the headword. The examples can either be authentic (exact quotations), adapted (modified versions of authentic examples) or invented examples ([?], p. 283-285). Authentic examples are attributed quotations (citations), which not only elucidate meaning and illustrate features of the headword (spelling, syntax, collocation, register etc.), but also function as attestations and are used provide evidence of the existence of a headword ([?], p. 453-458). We therefore call these examples “attestations”. Adapted examples, (cf a.o. [?], p. 251) or invented examples, which often occur in learner’s dictionaries and bilingual dictionaries ([?]), will not be discussed here.

2 Citations and attestations in scholarly dictionaries

In “scholarly dictionaries” (OED, ANW, the major historical dictionaries of Dutch¹, TLFi, Lewis and Short, Liddell and Scott, Böhrling and Roth, Hanyu Da Cidian, ...), the main role of citations and quotations is

1. To illustrate usage
2. To provide evidence for the lexicographical interpretation of a word sense (“attestation”)

As stated by Hawke, [?] p. 176, “The quotation evidence is the bedrock of any historical dictionary. The relationship between the definitions of each sense in a historical dictionary and the quotations that accompany it is particularly close. In the compilation of a historical dictionary (at least in modern times) the quotation evidence provides a sample of the empirical data on which the definitions have been based”

Thus, it is important to be explicit about the nature of the evidence provided. The dictionary should provide information on the reliability of attestations; this information should also be represented in the knowledge base.

2.1 Reliability of attestations

An attestation is as good as the corpus text or quotation the attestation is taken from. The analysis of the lexicographer as such can also be subject to discussion.

¹<http://gtb.ivdnt.org>

Uncertain reading or reconstruction

For historical dictionaries, the source material consists of text editions of historical documents. The further we go back in time, the more challenging it is to find enough material. Sometimes, sources have come down to us in one manuscript or a few manuscript fragments, with parts that are hard to read, or only partially legible (think of a hole in a manuscript). Or the word in question is misspelled.

The impact of this on the lexicographer's work is, that some of the "evidence" in a quotation is in fact a reconstruction².

Some examples In the Dictionary of Old Dutch, the entry with the Dutch adverb *fora* has related entries, one of which refers to a placename *foraskōta*. The editor suggests a correction of a spelling in the third quotation.

☞ **FORASKŌTA**^{II} *Voorschoot, Voorschoten, plaats bij Antwerpen, prov. Antwerpen Zie ook: skōta
Frescota. Voorschoten. TW 1025 [z.p.], 1144.
Verschote. Voorschoten. TW 1025 [z.p.], 1157.
Voscotle (l. vorscothe ?, red.). Voorschoten. TW 1025 Kamerijk, , 1181.

Figure 1: example: Spelling correction proposed by editor

Another example is from the Dictionary of Early Middle Dutch, where the fourth attestation has an occurrence of the verb *staen* (to stand), which is partly a reconstruction: s[tut].

☞ 12. Geneigd zijn tot.
ter quateden so staet hem hare moet Nat.Bl.D p. 76, r. 17, West-Vlaanderen, 1287
te wrake stuont her gere Dies rande manech man. di twie triland an. Trist. p. 341, r. 25-27, Nederrijn, 1250
Si wiste wel dat hem die moet Te wereliken saken stoet Lutg.K p. 212, r. 13-14, Brabant-West, 1265-1270
Ane sente Seruase s[tut] sin mut. [w]ant he dede [heme] [maneg] gut. Servas p. 296, r. 4-5, Limburg, 1200

Figure 2: example: reconstructed reading

These corrections or reconstructions are recognizable. According to the lexicographer, they are valid examples of the meaning of the entry. However, as representative of the spelling, they are not so reliable.

Uncertain interpretation

Sometimes, there may not be any doubt about the reading of a quoted passage of a source text, but the lexicographer is uncertain about his/her interpretation. In this example, the lexicographer gives a potential "misschien ook (maybe also)" other interpretation of an attestation.

*In onderstaande aanh. is behalve een acc.pl. **misschien ook** een interpretatie als znw.v. met als bet. 'bank(instelling)' mogelijk (vgl. MNW IV, 746-749).*

Oec mach die greue van vyanen lombarde houden binnen sinen sonderlinghen lande, *Corp.I* p. 2470, r. 20-21, Grimbergen, Brabant-West, 1298

Figure 3: example: doubt about interpretation

2.2 Citations as evidence: different types

There are several ways in which a citation as evidence can occur (list might not be exhaustive).

1. A bibliographical reference is given
2. The attested wordform is quoted in context + the bibliographical reference is given
3. Some context words are given, accompanied by a bibliographical reference.

²Another matter, which will briefly be discussed in section ??, is to be explicit about what the attestation actually provides evidence for in terms of the temporal and regional distribution of the phenomenon attested. For this purpose, the metadata provided with the quotations should accurately represent the nature of the source text from which the quotation has been taken.

Example of a and b In the Dictionary of the Dutch language, the quotation section begins with a listing of the dictionaries that describe the entry in the same meaning, followed by a group of quotations + bibliographical information with attestations (bold) of the entry in this quotation

2. (Bedr. en onz.) De feiten, de juiste toedracht, de waarheid kennen omtrent wat in het object of de voorzetselbep. wordt genoemd of bevraagd.
- a. Ter aanduiding van het bezitten van feitelijke kennis.
- ↪ α. Met een objectszin, ingeleid door een vragend vnw. of bijw.
- V. BERLAINMONT E iij v^o a [1536].
 LAMBRECHT, *Naembouck* [1546].
 PLANT. [1573].
 KIL. [1588].
 HEXHAM [1648].
 V. DALE [1872].
- Doe gingen de kinderen Juda tot Josua ... ende Chaleb ... sprac tot hem Ghi **weet** wat die Here tot Moyse den man Gods seyde van minen ende uwen wege in Kades Barnea, *Bijbel v. Liesveldt, Jozua* 14 B [1526].
 Noch langt my. en eewelijck sal om **weten** Wie sij waren wt goeder dueghden gront, J. V.D. DALE 173 [1528].
 Zouden zij **gheweten** hebben wat daer inne was ende bevindende dattet ghelt was, hebben tot mijns heere huuse ghebrocht, V. VAERNEWIJCK, *Ber. T.* 1, 174 [1566].
 Abraham berispte Abimelech ter oorsake eenes water-puts, die Abimelechs knechten met geweld genomen hadden. Doe seyde Abimelech; Ick en hebbe niet **geweten**, wie dit stuck gedaen heeft, *Statenb., Gen.* 21, 26 [ed. 1637].
 Alzoo (t.w. door middel van een paspoort) kan men **weeten** bijna, waar ieder zich ophoudt: alzoo is het wijduitgestrekt Rusland, voor de meerderheid zijner bewoners, eene soort van gevangenis, V. WOENSEL, *Rusl.* 193 [1804].
 En kijk hare dochteren maar eens aan, als gij **weten** wilt, hoe de deftige matrone er in den bloei van vrolijk- en jolijkheid uitzag, POTGIETER 1, 6 [1841].
 Ik **weet** nog niet waer de ongelukkige Geronimo verbleven is, CONSC., *S. Turchi* 2, 32 [1859].

An example of a and c In the Liddell Scott Jones lexicon, bibliographical references are given, or a description of the type of context the entry word was found in.

ἀνώμα^λ-ος, ov, (ἀ-priv., ὁμαλός)

A. *uneven, irregular*, “χώρα” **Pl.Lg.625d**; “φύσις” **Id.Tl.58a**; “τὸ ἀ. τῆς ναυμαχίας” **Th.7.71** (cj.), cf. *Arist.Pr.885a15*; and in *Sup., Hp.Aër.13*; of movements, *Arist.Ph.228b16*, al.; of periods of time, *Id.GA772b7*; of the voice, *ib.788a1*. Adv. “-λως, κινεῖσθαι” *Id.Ph.238a22*, cf. **Pl.Tl.52e**.

II. of conditions, fortune, and the like, “φεῦ τῶν βροτειῶν ὡς ἀ. τύχαι” *E.Fr.684*; πόλις, πολιτεία, **Pl.Lg.773b, Mx.238e**; “θεῶν” *Plot.6.7.34*. Adv. “-λως” **Hp.Prog.3, Isoc.7.29**; ἀ. διατεθῆναι τὸ σῶμα fall into *precarious* health, *Prisc.p.333 D*.

III. of persons, *inconsistent, capricious*, “ὁμαλῶς ἀ.” **Arist.Po.1454a26**; ὄχλος, δαιμόνιον, **App.BC3.42, Pun.59**; “πίθηκος” *Phryn. Com.20*; “τύχη” *AP10.96*. Adv. “-λως” **Isoc. 9.44**.

IV. Gramm., of words which deviate from a general rule, *anomalous*, *Diom.1.327 K.*; but τὸ ἀ. τῆς συντάξεως *diversity* of construction, *A.D.Synt.291.17*. Adv. “-λως” *Sch.Th.Qxy.853v18*.

2.3 Linked data modeling of lexical citations

The concept of attestations is discussed in two recent publications. One is the publication on the model for the DiaMaNT lexicon ([?]). The most recent publication is the one by Khan and Boschetti [?]. This article on lexical attestations prompted us to re-evaluate the proposal for attestations in Depuydt en Does 2018. We will discuss briefly the model by [?] and will then come up with an alternative proposal.

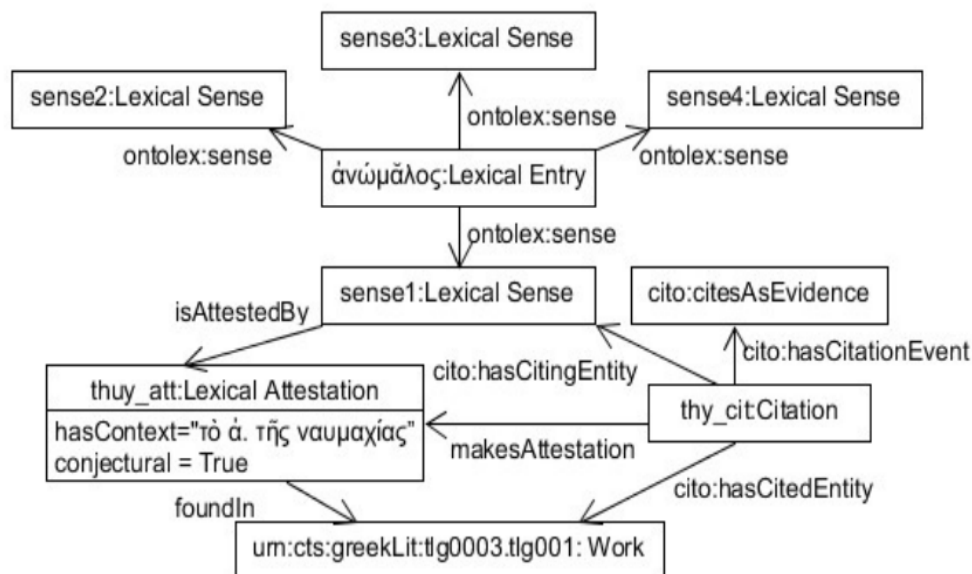


Figure 4: Example modeling from [?]

Khan and Boschetti’s *lemonBib* model for lexicographical citations tackles some issues important for historical lexicography and proposes solutions based on the FRBR, Cito and Fabio ontologies[?]

- Distinction between citation in general and a citation which provides evidence for a certain lexical object (*Attestation*) (Word sense, word form,)
- Enabling the marking of text readings as conjectural
- The distinction between a “work” and its “Manifestation” in the form of a publication.

In our view, the resulting model³ has some drawbacks.

- In the K&B model, the presence of context entails attestation (the *hasContext* data property has domain *Attestation*). The model does not take into account citations with context snippets that are *not* attestations, eg. citations in a section on etymology, where an authority is quoted to back up or contradict an analysis.
- There appears to be an unnecessary amount of relation-reifying objects in the model.

In the simplest application of the Cito model, “cites” is essentially a relation between two resources representing publications, without any intermediate reifying objects. Ported to the lexical domain, this would consist of a relation between a lexical sense and a work containing an occurrence of this sense. Since we want to attach some characterizations to the citation, some degree of reification is necessary, for which purpose Cito proposes the *Citation* class. K& B proceed by modeling *Citation* and *Attestation* as different resources with a relation *attestationCitation/makesAttestation* between them. We do not see why one could not declare *Attestation* as a subclass of *Citation* and characterize them, for instance, by (e.g.)

- *Attestation* is a subclass of *Citation*
- *Attestation* is a subclass of (*hasCitationCharacterization* cito:citesAsEvidence)

³<http://lari-datasets.ilc.cnr.it/lemonBib>