**Appendix: Data and queries**

The current version of the Late Latin Charter Treebank (LLCT) is available in .pml.xml format at the author’s Zenodo repository: https://doi.org/10.5281/zenodo.1197357. LLCT consists of documents from the following three copyright-free editions: *Codice diplomatico longobardo* (*CDL*) 1–2 (Luigi Schiaparelli, 1929–1933), *Codice diplomatico toscano*, part 2, vol. 1 (Filippo Brunetti, 1833), *Memorie e documenti per servire all’istoria del Ducato di Lucca* (*MED*), part 5, vol. 2 (Domenico Barsocchini, 1837).

The queries are written in Prague Markup Language Tree Query (PML-TQ) language for the homonymous extension in [TrEd Treebank Editor](https://ufal.mff.cuni.cz/tred/). LLCT is meant to be queried in TrEd with a customised version of the [ALDT schema](https://www.dropbox.com/s/x7cl64hjvvg817o/aldt_schema.xml?dl=0), which includes the LLCT-specific XML attributes, such as *seg* and *status*.

The formulaic/free distinction is encoded in LLCT with the *seg* (= segmentation) tag. The same attribute also indicates whether the tagged word is part of an autograph subscription (value *subs*). Since subscriptions are highly formulaic, they are here analysed together with formulaic words. The queries are written so that the output has to be copied to and manipulated in a spreadsheet. The table below presents the total word counts of the formulaic/free/subscription parts of LLCT to which the numbers of occurrence are compared. Sentence count cannot be used as a reference point because the sentence boundaries of documentary Latin are opaque and the ways to define them differ between editions.

**Table i.** The sizes of the textual segments in LLCT

|  |  |
| --- | --- |
| Segmentation | Number of words |
| formulaic | 156,032 |
| free | 56,314 |
| autograph subscription | 13,488 |

The *status* attribute indicates whether the inflexional ending of the word in question contains an expansion of an abbreviation (*expan*), a restoration of damaged letters (*damage*), or neither of them (*normal*). Only *normal* words can be utilised for queries that concern morphology. For a detailed account, see Korkiakangas & Passarotti (2011).

**Queries**

**1. Innovative lemmas**

# 81 innovative lemmas with segmentation status

aldt-sentence

[ descendant aldt-word $a :=

[ lemma in {"aldia1", "aldiaricus1", "aldio1", "aldionalis1", "arimannus1", "arra1", "banda1", "barba3", "barbanus1", "batto1", "bluto1", "bullitanus1", "calderarius1", "cambiator1", "cambium1", "caminata1", "canavarius1", "cavallarius1", "cavallicatura1", "cergiolitum1", "caesa1", "concambiatio1", "concambio1", "concambium1", "debluto1", "fiuwadia1", "focacia1", "fossata1", "fossatum1", "fumarius1", "gahagium1", "gasindus1", "gastaldus1", "grunda1", "launechild1", "mallo1", "marepas1", "marscalcus1", "monto1", "morgingabum1", "mustariolum1", "ornile1", "patrinius1", "paupertacula1", "petia1", "petiola1", "petiolum1", "petium1", "rasula1", "recta1", "rectum1", "scafilus1", "scherpha1", "scufia1", "sculdahis1", "spanga1", "sporus1", "staffilus1", "stanclio1", "stantarium1", "strata1", "summarra1", "sundrialis1", "sundrium1", "tessero1", "tia1", "tingatio1", "torta1", "ubiscarius1", "usitile1", "wadia1", "wadio1", "waldemannus1", "waldus1", "vassus1", "weregeldum1", "viccia1", "viganatio1", "viganio1", "vitellata1", "zapa1"} ] ];

>> give $a.lemma,$a.seg

**2. Future perfect form**

# future perfect form

aldt-sentence

[ descendant aldt-word $a :=

[ status = "normal", voice = "active", pos = "verb", tense = "future\_perfect" ] ];

>> give$a.form,$a.seg

**3. Dative plural form**

# dative plural form (3rd, 4th and 5th declension)

aldt-sentence

[ aldt-word

[ descendant aldt-word $a :=

[ status = "normal", number = "plural", case = "dative", declension in {"3", "4", "5"}, (pos = "noun" or pos = "adjective") ] ] ];

>> give $a.form,$a.seg

**4. Genitive vs. *de +* PP**

The genitives and the *de* + PPs are among the hardest-to-reach items in the present study. Ten queries are needed to cover all the relevant dependency structures, and after that manual disambiguation of the output is required. The figures illustrate the queried dependency structures.

**4.1** Genitives of type N-Ngen

N

|

Ngen

# N-Ngen: non-coordinated adnominal genitive (the head not headed by another nominal)

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ (pos = "noun" or pos = "pronoun"),

0x parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ],

0x parent aldt-word

[ pos = "conjunction",

parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ] ],

aldt-word $c :=

[ status = "normal", case = "genitive", (pos = "noun" or pos = "pronoun") ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$b.form,$c.form,$b.lemma,$c.lemma,$c.seg

**4.2** Genitives of type N-coord-Ngen

N

|

COORD

/ \

Ngen Ngen

# N-coord-Ngen: coordinated adnominal genitive (the head not headed by another nominal)

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ (pos = "noun" or pos = "pronoun"),

0x parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ],

0x parent aldt-word

[ pos = "conjunction",

parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ] ],

aldt-word $d :=

[ pos = "conjunction",

aldt-word $c :=

[ status = "normal", case = "genitive", (pos = "noun" or pos = "pronoun") ] ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$b.form,$c.form,$b.lemma,$c.lemma,$c.seg

**4.3** Genitives of type N-coord-N/Ngen

N

|

COORD

/ | \

N N Ngen

# N-coord-N/Ngen: adnominal genitive modifying coordinated Ns

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ relation in {"COORD", "APOS", "COORD\_AP"},

aldt-word $d :=

[ (pos = "noun" or pos = "pronoun"), (relation ~ "CO$" or relation ~ "AP$") ],

aldt-word $c :=

[ status = "normal", case = "genitive", !relation ~ "CO$", !relation ~ "AP$", !lemma = "is1", !lemma = "ipse1", !lemma = "qui1", (pos = "noun" or pos = "pronoun") ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$d.form,$c.form,$d.lemma,$c.lemma,$c.id,$c.seg

**4.4** Genitives of type N-Ngen-Ngen

N

|

Ngen

|

Ngen

# N-Ngen-Ngen: non-coordinated adnominal genitive with the head headed by another nominal (two-genitive chain of type rector ecclesie Stefani)

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ (pos = "noun" or pos = "pronoun"),

0x parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ],

0x parent aldt-word

[ pos = "conjunction",

parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ] ],

aldt-word $c :=

[ status = "normal", case = "genitive", (pos = "noun" or pos = "pronoun"),

aldt-word $d :=

[ status = "normal", case = "genitive", (pos = "noun" or pos = "pronoun") ] ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$c.form,$d.form,$c.lemma,$d.lemma,$d.seg

**4.5** Genitives of type N-Ngen-Ngen-Ngen

N

|

Ngen

|

Ngen

|

Ngen

# N-Ngen-Ngen-Ngen: non-coordinated adnominal genitive with the head headed by another nominal (three-genitive chain of type manus rectoris ecclesie Stefani)

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ (pos = "noun" or pos = "pronoun"),

0x parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ],

0x parent aldt-word

[ pos = "conjunction",

parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ] ],

aldt-word $c :=

[ case = "genitive", (pos = "noun" or pos = "pronoun"),

aldt-word $d :=

[ status = "normal", case = "genitive", (pos = "noun" or pos = "pronoun"),

aldt-word $e :=

[ status = "normal", case = "genitive", (pos = "noun" or pos = "pronoun") ] ] ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$d.form,$e.form,$d.lemma,$e.lemma,$e.seg

**4.6** Genitives of type N-Ngen-Ngen-Ngen-Ngen

N

|

Ngen

|

Ngen

|

Ngen

|

Ngen

# N-Ngen-Ngen-Ngen-Ngen: non-coordinated adnominal genitive with the head headed by another nominal (four-genitive chain)

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ (pos = "noun" or pos = "pronoun"),

0x parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ],

0x parent aldt-word

[ pos = "conjunction",

parent aldt-word

[ (pos = "noun" or pos = "adjective" or pos = "pronoun") ] ],

aldt-word $c :=

[ case = "genitive", (pos = "noun" or pos = "pronoun"),

aldt-word $d :=

[ case = "genitive", (pos = "noun" or pos = "pronoun"),

aldt-word $e :=

[ status = "normal", case = "genitive", (pos = "noun" or pos = "pronoun"),

aldt-word $f :=

[ status = "normal", case = "genitive", (pos = "noun" or pos = "pronoun") ] ] ] ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$e.form,$f.form,$e.lemma,$f.lemma,$f.seg

**4.7** PPs of type N-de-N

N

|

de

|

N

# PPs of type N-de-N

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ status = "normal", (pos = "noun" or pos = "pronoun" or pos = "adjective"),

aldt-word $c :=

[ lemma = "de1",

aldt-word $d :=

[ (pos = "noun" or pos = "pronoun" or pos = "adjective") ] ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$b.form,$c.form,$d.form,$b.lemma,$d.lemma,$b.seg

**4.8** PPs of type N-de-coord-N

N

|

de

|

COORD

/ \

N N

# PPs of typeN-de-coord-N: de + PPs with coordinated complements

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ status = "normal", (pos = "noun" or pos = "pronoun" or pos = "adjective"),

aldt-word $c :=

[ lemma = "de1",

aldt-word $e :=

[ pos = "conjunction",

aldt-word $d :=

[ (pos = "noun" or pos = "pronoun" or pos = "adjective") ] ] ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$b.form,$c.form,$d.form,$b.lemma,$d.lemma,$b.seg

**4.9** PPs of type N-coord-de-N

N

|

COORD

/ \

N de

|

N

# N-coord-de-N: de + PPs coordinated with genitives

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ status = "normal", (pos = "noun" or pos = "pronoun" or pos = "adjective"),

aldt-word $e :=

[ pos = "conjunction",

aldt-word $c :=

[ lemma = "de1",

aldt-word $d :=

[ (pos = "noun" or pos = "pronoun" or pos = "adjective") ] ] ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$b.form,$c.form,$d.form,$b.lemma,$d.lemma,$b.seg

**4.10** PPs of type N-coord-N/de-N

N

|

COORD

/ | \

N N de

|

N

# N-coord-N/de-N: de + PPs modifying coordinated Ns

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ relation in {"COORD", "APOS", "COORD\_AP"},

aldt-word $e :=

[ lemma = "de1",

aldt-word $c :=

[ status = "normal", !relation ~ "CO$", !relation ~ "AP$", (pos = "noun" or pos = "pronoun") ] ],

aldt-word $d :=

[ (pos = "noun" or pos = "pronoun"), (relation ~ "CO$" or relation ~ "AP$") ] ] ];

>> give $a.document\_id,$a.subdoc,$a.id,$d.form,$e.form,$c.form,$d.lemma,$c.lemma,$c.seg,$c.id

**Post-processing:** The queries provide the sentence ID numbers, so that the output sentences can be analysed in their full-text context when post-processing them. Since the annotation style of LLCT does not allow distinguishing between a genuine genitive construction (e.g., *manus presbiteri* “the priest’s hand”) and an attribute chain (e.g., *Teutperti presbiteri* “of the priest Teutpertus”, where both the genitives modify a common head, such as *signum* “cross”), the phrases of this type have to be disambiguated manually. A large group to be disambiguated are the NPs with *ipse*, *ille* and *hic*, where the pronoun can be either pronominal or adjectival: *ecclesie ipsius* “of his church” is a genuine possessive construction while *ecclesie ipsius* “of that church” is not.

**5. Accusative and infinitive vs. complement clauses**

**5.1** Accusative and infinitive with non-coordinated subject

# ACI with non-coordinated subject

aldt-sentence

[ descendant aldt-word $a :=

[ status = "normal", mood = "infinitive", pos = "verb",

aldt-word $b :=

[ relation = "SBJ" ] ] ];

>> give $a.form,$b.form,$b.seg

**5.2** Accusative and infinitive with coordinated subject

# ACI with coordinated subjects

aldt-sentence

[ descendant aldt-word $a :=

[ status = "normal", mood = "infinitive", pos = "verb",

aldt-word

[ (relation = "COORD" or relation = "APOS"),

aldt-word $b :=

[ (relation = "SBJ\_CO" or relation = "SBJ\_AP") ] ] ] ];

>> give $a.form,$b.form,$b.seg

**5.3** Phrasal complementation

# clausal object complements with conjunction (instead of ACI)

aldt-sentence

[ descendant aldt-word $a :=

[ pos = "verb", !mood = "infinitive",

aldt-word $b :=

[ relation = "AuxC", lemma != "si1",

aldt-word $c :=

[ pos = "verb", relation = "OBJ", !mood = "infinitive" ] ] ] ];

>> give $a.lemma,$a.form,$b.form,$c.form,$b.seg

**Post-processing:** The complementiser clauses were disambiguated in their full-text context to isolate subordinate interrogative clauses and comparative clauses, which cannot be excluded on the basis of the treebank annotation. Six cases were discarded (*facias comodo placueret* (*CDL* 51), *cogitare qualiter redimere* (*CDL* 171), *iscimus qualiter occhurra* (*CDL* 230), *iscio qualiter hoccurra* (*CDT* 12), *considerastis qualiter perveniret* (*MED* 523), *dicimus quod fiat* (*CDL* 68)).

**6. Absolute constructions**

**6.1** Absolute constructions with non-coordinated subject

# absolute constructions with non-coordinated subject

aldt-sentence

[ descendant aldt-word $a :=

[ pos = "participle", !mood = "gerundive", (relation = "ADV" or relation = "ADV\_CO" or relation = "ADV\_AP"),

0x aldt-word

[ relation = "AuxV" ],

aldt-word $b :=

[ relation = "SBJ" ] ] ];

>> give $a.form,$b.form,$a.seg

**6.2** Absolute constructions with coordinated subjects

# absolute constructions with coordinated subjects

aldt-sentence

[ descendant aldt-word $a :=

[ pos = "participle", !mood = "gerundive", (relation = "ADV" or relation = "ADV\_CO" or relation = "ADV\_AP"),

0x aldt-word

[ relation = "AuxV" ],

aldt-word

[ (relation = "COORD" or relation = "APOS"),

aldt-word $b :=

[ (relation = "SBJ\_CO" or relation = "SBJ\_AP") ] ] ] ];

>> give $a.form,$b.form,$a.seg

**7. Second-person singular form**

**7.1** Second-person singular form in -s

# active second person singular -s, non-perfect

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ status = "normal", person = "second\_person", tense != "perfect", number = "singular", voice = "active", mood != "imperative", form ~ "s$" ] ];

>> give $a.document\_id,$a.subdoc,$a.date,$b.form,$b.lemma,$b.seg

**7.2** Second-person singular form not ending in -s

# active second person singular non-s, non-perfect

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ status = "normal", person = "second\_person", tense != "perfect", number = "singular", voice = "active", mood != "imperative", form !~ "s$" ] ];

>> give $a.document\_id,$a.subdoc,$a.date,$b.form,$b.lemma,$b.seg

**8. Dative singular form**

# dative singular forms

aldt-sentence

[ descendant aldt-word $a :=

[ status = "normal", case = "dative", number = "singular", !declension = "0" ] ];

>> give$a.form,$a.seg

**9. Subject case encoding**

# non-coordinated non-pronominal imparisyllabic 3rd-declension singular subjects

aldt-sentence $a :=

[ descendant aldt-word

[ mood !~ "infinitive", !(pos = "participle" and tense = "present"), !(pos = "participle" and case = "ablative"), (pos = "verb" or pos = "participle"),

aldt-word $b :=

[ status = "normal", declension = "3", number = "singular", relation ~ "^SBJ", gender != "neuter", pos != "pronoun", animacy in {"-", "0"} ] ] ];

>> give $a.document\_id,$a.subdoc,$a.date,$b.form,$b.lemma,$b.seg,$b.case

# coordinated non-pronominal imparisyllabic 3rd-declension singular subjects

aldt-sentence $a :=

[ descendant aldt-word

[ mood !~ "infinitive", !(pos = "participle" and tense = "present"), !(pos = "participle" and case = "ablative"), (pos = "verb" or pos = "participle"),

aldt-word

[ (relation = "COORD" or relation = "APOS"),

aldt-word $b :=

[ status = "normal", declension = "3", number = "singular", relation ~ "^SBJ", gender != "neuter", pos != "pronoun", animacy in {"-", "0"} ] ] ] ];

>> give $a.document\_id,$a.subdoc,$a.date,$b.form,$b.lemma,$b.seg,$b.case

**10. Verb/object order**

**10.1** Verb/object order, formulaic parts

# VO/OV order with finite V, formulaic parts, only non-coordinated arguments

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ mood !~ "infinitive", !(pos = "participle" and tense = "present"), !(pos = "participle" and case = "ablative"), (pos = "verb" or pos = "participle"),

aldt-word $d :=

[ relation = "OBJ", iobj != "1", seg = "formulaic", (pos = "noun" or pos = "adjective") ] ] ];

>> give distinct $a.id,$b.form,$d.form,$b.id,$d.id

**10.2** Verb/object order, free parts

# VO/OV order with finite V, free parts, only non-coordinated arguments

aldt-sentence $a :=

[ descendant aldt-word $b :=

[ mood !~ "infinitive", !(pos = "participle" and tense = "present"), !(pos = "participle" and case = "ablative"), (pos = "verb" or pos = "participle"),

aldt-word $d :=

[ relation = "OBJ", iobj != "1", seg = "free", (pos = "noun" or pos = "adjective") ] ] ];

>> give distinct $a.id,$b.form,$d.form,$b.id,$d.id

**Post-processing:** The query outputs the sentence identifier numbers of V and O which indicate the linear position in the sentence. The relative order of the arguments must be defined on the basis of these numbers.