

# Digital Aquinas

## NLP support for study and translation of Latin texts

Martin Holub, Benedikt Mohelník, Pavel Straňák, Dušan Variš, Anna Vernerová, Šárka Zikánová

# Digital Aquinas

- Summa Theologiæ of Thomas Aquinas
  - Latin original written 1265–1274, cca 3500 pages,
  - first digital NLP corpus: Index Thomisticus (Busa, IBM 1949–)
  - crucial work of European philosophy, many translations
- Facilitating research (philosophy, history, translation, ...)
  - NLP processing: segmentation, tokenisation, morphology and lemmatisation (medieval Latin) (+maybe more)
  - Latin, Czech, English (Polish, ...) >27000 alignment pairs in Kontext (~ paragraphs)
  - Treq (or similar interface), terminology extraction, checking consistency of translation
  - word alignment in Kontext (how?), reading interface with rich information (planning)



# Interesting Aspects

- Medieval Latin processing
  - even though TA is said to be quite good, “engineers’ latin” in this aspect
- Czech translation from 1930’s is extremely literal
  - good for alignment, bad for morphology (invented words), interesting from the translation aspect
- English translation much less literal, new Czech one too. Doesn’t mean they resign on accuracy, just try to reach it differently.
- Alignment (not necessarily the same as for MT)
- More processing like detecting and linking citations (in ST Aristotle and others)



# Data

	Latin	Czech
positions	1,897,306	2,061,314
Attributes		Attributes
word	49,547	56,227
lc	46,808	51,068
Structures		Structures
<doc>	1	1
<align>	91,275	91,275
<div>	26,513	26,513
<p>	24,796	29,261
<i>	11,660	5



# 2 projects submitted:

- **GACR project** of *Martin Holub and the team of translators*: linguistic aspects of the translation (medieval latin, theological terminology)
- NAKI project: **CREANT – Aplikace pro studijní čtení a analýzu českých překladů**
  - **Kontext + TEI:TOK + dict. apps** > some new stuff
  - platform for humanities corpora
  - Straňák, Holub, Janssen, Svitalský, Šimandl, Klímová, Mohelník ...



# NLP requests in Humanities

- mostly in LINDAT project, but also from outside
- modern history: state archives for historians – land records, etc. Czech, Slovak, forms, typewriters, hand...
- OCR for massive corpus of Czech: libraries digitization
  - all domains including very technical, all ages
  - get full metadata (title, author, italics, image+caption...)
- corpora of Patočka, Komenský, Diviš:
  - works, translations, letters
  - facsimiles, digitisation, rich corpora, study interface



principale ad recedendum a peccato et resistendum concupiscentiis eius , quae in contrarium caritatis movent . Et hoc pertinet ad incipientes , in quibus caritas est nutrienda vel fovenda ne corrumpatur . Secundum autem studium succedit , ut homo principaliter intendat ad hoc quod in bono proficiat . Et hoc studium pertinet ad proficientes , qui ad hoc principaliter intendunt ut in eis caritas per augmentum roboretur . Tertium autem studium est ut homo ad hoc principaliter intendat ut Deo inhaereat et eo fruatur . Et hoc pertinet ad perfectos , qui cupiunt dissolvi et esse cum Christo . Sicut etiam videmus in motu corporali quod primum est recessus a termino ; secundum autem est appropinquatio ad alium **terminum** ; tertium autem quies in termino .

Ad primum ergo dicendum quod praemium est **terminus** meriti . Est autem duplex terminus motus , scilicet ultimus ; et medius , qui est et principium et terminus . Et talis terminus est merces augmenti . Merces autem favoris humani est sicut ultimus terminus his qui finem in hoc constituunt , unde tales nullam aliam mercedem recipiunt .

@I-II q. 114 a. 8 ad 1

odpírati jeho žádostem , jež pudí k opaku lásky . A to náleží začátečníkům , v nichž má být láska živena a opatrována , aby se nezrušila . Nastane pak druhá snaha , by člověk hlavně dbal toho , by v dobrém prospíval . A tato snaha náleží prospívajícím , kteří dbají hlavně toho , aby se v nich láska zvětšením upevnila . Třetí pak snaha jest , aby dbal člověk hlavně toho , by přilnul k Bohu a jeho požíval . A to náleží dokonalým , kteří " touží se vyprostiti a býti s Kristem " . - Jakož také vidíme při pohybu tělesném , jenž nejprve je vychod od konce , druhé je pak přiblížení k jinému konci , třetí jest klid v konci .

K prvnímu se tedy musí říci že odměna je cílem zásluhy . Jest však dvojí cíl pohybu ; totiž poslední a střední , jenž jest i původ i cíl ; a takovým cílem je mzda zvětšení . Avšak mzda lidské přízně jest jako poslední cíl těm , kteří do toho kladou cíl ; pročež takoví nedostanou žádné jiné mzdy .

K prvnímu se tedy musí říci , že onen důvod má dvě chyby v projednávaném . A to nejprve , protože

# Kontext

## study of translations of ‘terminus’ in Summa Theologiae

INT - As suas razões / e os seus interesses para vir para Portugal e quais são os seus interesses na língua na cultura <portuguesa> / portuguesas o que é que acha de Portugal dos Portugueses //

INF - Ok então eu *Ah* estou aqui em Portugal <porque> / por vários motivos um porque Itália não se sabe muito de Portugal //

INF - *Eh* não sei falar muito de Portugal *Ah* eu acho que muita pessoa *Ah* poder pensar que o Portugal *es Espanha* //

INF - *Ah* em Natal quando eu fui em casa / *Eh* meu tio [me] [me] me perguntou então como é a Espanha? *Eh hhh* e aí explicar *no no* é assim *Bater de mesa* //

INF - E já por isso pa conhecer <uma> um estado uma cultura uma população que / *n@* não conheço não *s@* / ante de ser aqui não sabia nada do Portugal //

INF - Depois porque eu estudo Geografia / *Eh Eh* a escola portuguesa de Geografia é uma escola muito importante *en Europa* / então / achavo e acho que podia ser [uma] uma boa escolha pa [a] a minha / *carrei@* carreira universitária não sei como se diz carreira //

INF - *Eh* sim [é] é isso é tudo essa motivação principal [porque] / porque estou aqui estou aqui *da setembro* *ma* o meu Erasmus se iniciou *en janeiro* //

INF - *Eh* *ma* //

INF - Aproveitei pra estudar um pouco mais português porque em seis meses *Eh* *[n@]* *no* consigo [e] *de@ Uhm* e pra iniciar a trabalhar sobre [a] a tese / *Eh* depois da cultura portuguesa *Eh Uhm* //

INF - A cultura portuguesa que eu conheço [é] é *Uhm* concentrada no âmbito [da] [da] da minha *m@* disciplina da Geografia então [é] é colegada de exploração *Ah Uhm* o viagem do Portugal e de como os portugueses difunderam a própria cultura <no> no vários paíse //

INF - *Eh Bater de mesa* é só aquelo que o co@ <conheço> co@ conheci agora *Ah* já aproveito pra ver pra conhecer *Eh* tudo o Portugal a próxima semana *va@* vou à *Uhm* aldeia do xisto pra ver [como] como é porque conheço solo *Eh* *ma* nunca fui //

INT - Muito bem e o que é que o surpreendeu mais no contacto com os portugueses? / Acerca do povo português o que é que / o surpreendeu? //

INF - *Eh Uhm* // [uma] [uma] [cosa] uma cosa estranha porque no *Uhm* / eu achavo que *Eh n@ no Uhm* / *no* sei explicar / *Ah* muita vezes <no> / *Ah* encontro portu@ *no* sei se é solo *Ah Uhm* / [a] a pessoa mais nova porque con <a> *Uhm* / [...] falo con a perso@ pessoas veria também e não são assim *ma Ah* [a] a pessoa mais nova o minha colegas da universidade son pessoa que muita vezes são fechados / e não sei porquê porque solitamente os povo <de> *Uhm* / [do] do litoral sul de Europa *[n@]* *no* são assim / *m@* *ma* estranho porque ap@ conheço os pai [da] <minha companheira> da *Uhm Ah* minhas companheiras de casa [e] [e] e *no* são assim mas conheço / colega de universidade que muita vezes vou a cumprimentar-me / ele nunca vão a responder entã@ *hhh hhh* o que *hhh* não sei se //

INT - *hhh* //

INF - *hhh* só isso //

INT - [E] E qua@ quais são as diferenças ou as principais diferenças ou semelhanças que encontra entre os portugueses e os italianos? //

TEI:TOK  
rich representation of documents



EN | CZ

Old Czech TEITOK Kontext

- [Home](#)
- [Documents](#)
- [CQP Search](#)
- [Manatee Search](#)
- [Kontext](#)
- [Login](#)

Powered by TEITOK  
© Maarten Janssen, 2014-

## Corpus Search

CQP Query: [word="zlom"]

[Search](#) [query builder](#)

2 results

Tags: [POS tag \(non-disambiguated\)](#) [Lemma](#)

context Vezmi chléb horký žitný , **zlom** a oblož [...] hlavy všady ,

context Vezmi chléb horký žitný, **zlom** a oblož ... hlavy

[Download results](#) - Remember query

### Frequency Options

Collocation by: [\[select\]](#) | Context size: [1](#) | Direction: [Left and Right](#) [Submit](#)

Frequency by: [\[select\]](#)

TEI:TOK  
corpus search (or one can use Kontext)

## [Knihy lékařské]

Title [Knihy lékařské]

### View options

Show: - Tags:  POS tag (non-disambiguated)  Lemma

index

Folio 1.0

> 1v

[1]

... s tiem polejем a daj mu jiesti [...] k rozumu a k zdraví .

Opět k témuž jinak

Vezmi chléb horký žitný , zlom a oblož [...] hlavy všady , i židoviny a potom [...] lehnúti . A to učiň dvakrát [...]

Komu mozk zschne v hlavě

Vezmi hrnec a zvrť dierky v [...] múky z jarého žita <sup>[7]</sup> dobře a vstav po dvakrát [...] A potom tu múku vař [...] do téhotné po málu .

Komuž mozk zschne , že hlava [...]

Pomni a měj toto naučenie , vezma [...] múky , vespi v nový hrnec [...] v pec , když chléb pekú [...] zberiž na lžičku tu múku , peka [...] a s kmínem <sup>[a]</sup> lučným a [...] ořechy , rozetra to v [...] múkú . A potom beřiž [...] a učiň sobě [...] polež [...] semenec [...]

[Download XML](#) • [Download text](#)

# TEI:TOK

search results in full documents

[1]

... s tiem polejem a daj mu jiesti ...<sup>[2]</sup> k rozumu a k zdraví.

Opět k témuž jinak

Vezmi chléb horký žitný, zlom a oblož ...<sup>[3]</sup> hlavy všady, i židoviny a potom ...<sup>[4]</sup> lehnúti.

A to učiň dvakrát ...<sup>[5]</sup>

Komu mozk zschne v hlavě

Vezmi hrnec a zvrtaj dierky v ...<sup>[6]</sup> múky z jarého žita<sup>[7]</sup> dobře a vstav po dvakrát ...<sup>[8]</sup> A potom tu múku vař ...<sup>[9]</sup> do téhodne po málu.

Komuž mozk zschne, že hlava ...<sup>[10]</sup>

Pomni a měj toto naučenie, vezma ...<sup>[11]</sup> múky, vespi v nový hrnec ...<sup>[12]</sup> v pec, když chléb pekú ...<sup>[13]</sup> zbeřiž na lžičku tu múku, peka ...<sup>[14]</sup> a s kmijenem lučným a ...<sup>[15]</sup> ořechy, rozetra to v ...<sup>[16]</sup> mukú. A potom beřiž ...<sup>[17]</sup> a učiň sobě ...<sup>[18]</sup> polež ...<sup>[19]</sup> semenec ...<sup>[20]</sup> ...

