

# The role of CLARIN and linguistic annotation in the digital transformation in research



**Martin Wynne**

[martin.wynne@bodleian.ox.ac.uk](mailto:martin.wynne@bodleian.ox.ac.uk)



Bodleian Libraries &  
Faculty of Linguistics, Philology and Phonetics,  
University of Oxford

National Coordinator, CLARIN-UK

KB } national library  
of the netherlands

***Libraries as Research Partners in  
Digital Humanities***

DH2019 Pre-conference, The Hague

Monday 8<sup>th</sup> July 2019

# Oxford Text Archive, 43 years on and in a new home



DIGITAL BODLEIAN

DIGITAL BODLEIAN MAKES THESE EXTRAORDINARY LIBRARY COLLECTIONS AVAILABLE ONLINE FOR THE VERY FIRST TIME...

Enter new search term   [SEE HOW IT WORKS](#)

THE COLLECTIONS

WESTERN MANUSCRIPTS

ORIENTAL

Oxford Dictionary of National Biography

Henry Oldenburg (1618-1677), scientific correspondent and secretary of the Royal Society, was born in the free city of Bremen, Germany, one of two children of Heinrich Oldenburg (d. 1634), a member of the Bremen Pseudepigram and later a professor at Dorpat University. His ancestors, originally from Meissen, had been Catechist teachers since the mid-seventeenth century. Oldenburg's grandfather, also Heinrich, had acquired the building of the vicarage of St. Lorenz at the present cathedral, in return for which he, his son, and his grandson inherited an annual income of 180 rixsdollars, which was paid only fitfully and precariously.

**Developing years**

Oldenburg began his education at the Pseudepigram where his father taught, and transferred in May 1633 to the Gymnasium illustre for a thorough grounding in religion, classical languages, the liberal arts, and theology; he received the degree of master of theology on 2 November 1639 for a thesis on relations between church and state. In May 1641 he left Bremen for the Netherlands with letters of introduction to various scholars. (He was perhaps influenced by the fact that his sister was about to marry Heinrich Koch, a relation of Johannes Koch, or Casparus, professor of theology at Leiden.) Oldenburg settled briefly in Utrecht, which possessed a lively university where Cartesianism was taught. However, he found Utrecht expensive and began to aspire to look for employment 'instructing either the son of a substantial or the son of some honest merchant' (Correspondence, 1.3-4). For the next twelve years he travelled on the continent and in England as tutor to a succession of young men, mostly English, with both royalist and parliamentary connections, as indicated by surviving drafts of letters in his correspondence book. His aim was not only to earn his living but to learn about relations between church and state and to see the world. In all of this he was clearly very successful, becoming fluent in the languages of the countries he visited. He met many Englishmen living abroad, notably John Donne and John Wall, with both of whom he later had connections.



## Networking the Republic of Letters, 1550-1750

Established in 2009, and now in our fourth phase (April 2017 - September 2018), we are a collaborative, interdisciplinary research project based at the University of Oxford with funding from The Andrew W. Mellon Foundation. We are using digital methods to reassemble and interpret the correspondence networks of the early modern period.



The Letter Catchers by Annaliese Stony for Cultures of Knowledge (with inspiration from Pier Francesco Molà)

electronic enlightenment

what's new | about ee | content | collaborate | publish | subscriber services | contact us | logout

INFORMATION | PRINT-HOUSE | COFFEE-HOUSE | BROWSE | SEARCH | HELP

Logout

INFORMATION

ABOUT EE

CONTENT

COLLABORATE

PUBLISH

SUBSCRIBER SERVICES

CONTACT US

WELCOME TO EE


"The useful information and experience you would have opportunities of acquiring are greater than you can conceive..." — Sir Samuel Benham to Jeremy Benham (11-12 April 1780)

Your access is provided by: University of Oxford. Served by ee-cloud-x


Current release: Autumn 2016

Latest news from Electronic Enlightenment . . .

**Autumn 2016 update!**

 Our Autumn 2016 update adds over 500 documents, adding especially to the Bernardin de Saint-Pierre edition. — Read all about it in the Autumn 2016 News-sheet.

**Voltaire's love letters**

 In "Voltaire's love letters" Amela Carruthers reminds us that philosophers can also be lovers! — Read all about it on the OUPblog


**Anniversarial for 2017**

**CALENDAR**

10 fall of clock 11  
12 Apogee. 12  
13 rise 7 h. 26 m  
14 7th path Iria. 8  
15 may back for  
16 Daybreak 4 h. 37

There are many correspondents in *EE* whose anniversary of birth or death is marked in 2017. From the 400<sup>th</sup> anniversary of the birth of the physician Benjamin Worsley (born 1617–died 1677) to the 200<sup>th</sup> anniversary of the death of salon hostess and author Anne Louise Germaine de Staël, Baroness Staël von Holstein (born 1766–died 1817). — Visit the 2017 Anniversarial.

**Spring 2016 Miscellany!**

 In "Enlightenment & Visual Impairment in the 17th & 18th Centuries" Jack Orchard examines attitudes to blindness. — Read all about it in the Spring 2016 Miscellany.



THE NEWTON PROJECT

Home The Texts About Newton About Us Support Us Search

Welcome to the Newton Project

The Newton Project is a non-profit organization dedicated to publishing in full an online edition of all of Sir Isaac Newton's (1642-1727) writings — whether they were printed or not. The edition presents a full (diplomatic) rendition featuring all the amendments Newton made to his own texts or a more readable (normalised) version. We also make available translations of his most important Latin religious texts.

Although Newton is best known for his theory of universal gravitation and discovery of calculus, his interests were much broader than is usually appreciated. In addition to his celebrated scientific and mathematical writings, Newton also wrote many alchemical and religious texts and he left many administrative papers in his role as Warden and then Master of the Mint.



© 2017 The Newton Project

Professor Rob Eiffe  
Director, AHRC Newton Papers Project

Scott Mandelstam,  
Fellow & Former Librarian, RevHouse, Cambridge

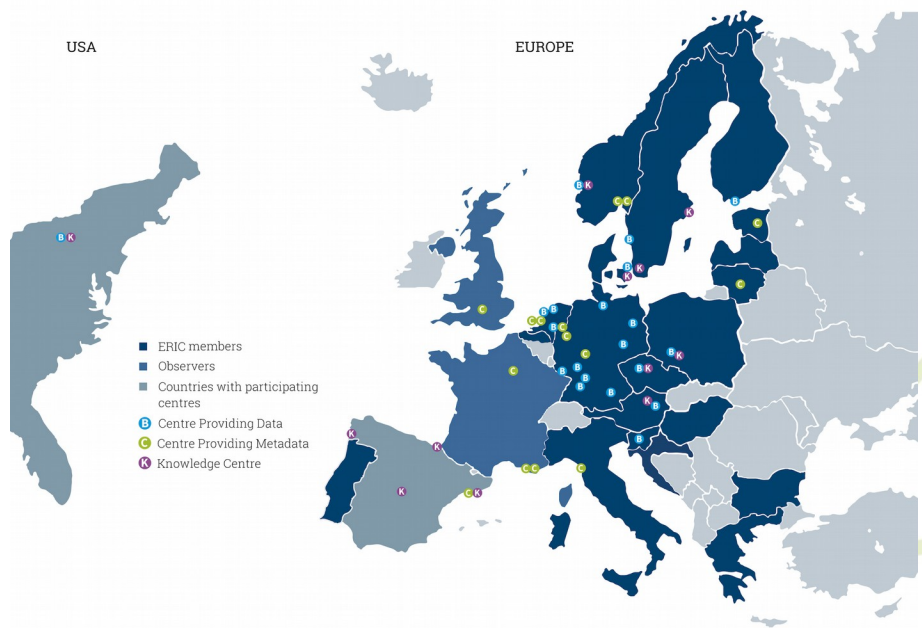
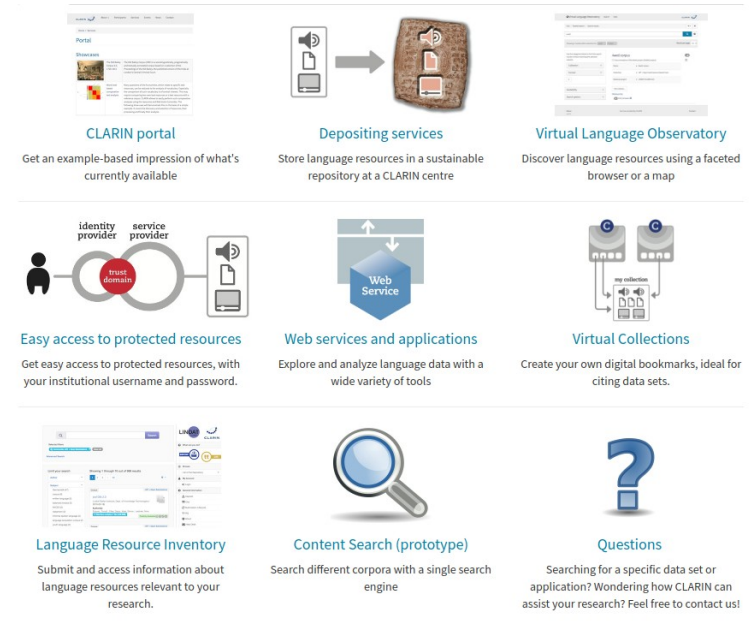
Faculty of History, George Street, Oxford, OX1 2PL -  
newtonproject@history.ox.ac.uk

Privacy Statement



# CLARIN in a nutshell

- **CLARIN** is the Common Language Resources and Technology Infrastructure;
- it provides easy and sustainable access for scholars in the **humanities and social sciences** and beyond
- to **digital language data** (in written, spoken, video or multimodal form),
- and **advanced tools** to discover, explore, exploit, annotate, analyse or combine them, wherever they are located
- through a **single sign-on** online environment.



## What's it all about?

- Discovery
- Support
- Sustainability
- Interoperability
- Cross-border collaboration



# *The datasets*

**Electronic Enlightenment:** a collection of scholarly digital editions of correspondence.

Collection started with Voltaire & Rousseau, expanded into other Enlightenment thinkers and writers, and has expanded into other eras, languages and domains. Now 80 000+ letters, 15 000+ people

A project in the University of Oxford, initially part of the Voltaire Foundation, moved to the Bodleian Libraries in 2008. Access is via institutional subscription managed by Oxford University Press.

**Oxford Text Archive:** 63000+ historical texts, language corpora, and other outputs of digital humanities research project; mostly open access, but some resources restricted to authenticated academic users.



Logout

immensely  
resource for  
2017 Anniversary  
A listing of some of those  
correspondents whose  
birth or death is marked  
in 2017. — BSECS

... read it now!



Bodleian Libraries  
UNIVERSITY OF OXFORD



Your access is provided by:  
University of Oxford. Served by  
ee-cloud-c

Current release: Autumn 2016

## Electronic Enlightenment — letters & lives online

... reconnecting the first global social network!

*Electronic Enlightenment* is the most wide-ranging online collection of edited correspondence of the early modern period, linking people across Europe, the Americas and Asia from the early 17th to the mid-19th century — reconstructing one of the world's great historical "conversations".

A **subscription** to *EE* will give you immediate access to **70,057** historical documents. Listen in on the first global social network as **8,560** historical figures discuss everything from religious tolerance to animal rights, vulcanology to classical archeology, economic modelling to celebrity culture.

---

### EE update Autumn 2016

---

#### *Electronic Enlightenment Scholarly Edition of Correspondence*

Now over 70,000 items of edited correspondence! Our latest update adds over **500 documents**, adding especially to the Bernardin de Saint-Pierre edition.

#### *Electronic Enlightenment Biographical Dictionary*

With over 100 additions and revisions our Biographical Dictionary now has over 8,500 entries!

For an overview of the update please read the latest **News-sheet**.



Not sure where to start? Then why not:

- take a **Guided tour**;
- **Browse** the list of correspondents;
- read our latest **News-sheet**;
- brush up on our **Fact sheet**.



Duell-ease A worde with. valiant spiritts shewing the abuse of duells, that valour, refuseth challenges and priuate combates. sett foorth by G.F. a defendour of Christian valoure.

“ Please use the following text to cite this item or export to a predefined format: BIBTEX CMDI

G. F., defendour of Christian valoure., 2013, *Duell-ease A worde with. valiant spiritts shewing the abuse of duells, that valour, refuseth challenges and priuate combates. sett foorth by G.F. a defendour of Christian valoure.*, Oxford Text Archive, <http://hdl.handle.net/20.500.12024/13791>

Share:

[Oxford Text Archive](#)

Authors	G. F., defendour of Christian valoure.
Date issued	2013-12
Type	Text
Language(s)	English
Publisher	University of Oxford
Collection(s)	Early English Books Online (Phase 2)

[Show full item record](#)

Files in this item [Download all files in item \(428.28 KB\)](#)

This item is **Academic Use** and licensed under: Text Creation Partnership

<b>Name</b>	A00503.epub
<b>Size</b>	73.94 KB
<b>Format</b>	Unknown
<b>Description</b>	Version of the work for e-book readers in the EPUB format

[Download file](#)

<b>Name</b>	A00503.html
<b>Size</b>	168.64 KB
<b>Format</b>	HTML
<b>Description</b>	Version of the work for web browsers

[Download file](#) [Preview](#)

<b>Name</b>	A00503.xml
<b>Size</b>	174.1 KB

What can you do?

[DEPOSIT](#) [CITE](#)

Browse

> All of the Repository

My Account

Login

Statistics

Statistics **BETA**

General Information

Deposit

Cite

Submission Lifecycle

FAQ

About

Help Desk

# What are we ultimately aiming for?

Ways to combine *close reading* with *big data* approaches.

# *Close reading v. distant reading*

"It is not easy to justify assertions about the alleged frequency of infrequency of some particular belief or attitude in the past. How many examples does one need to cite in order to prove the point? Lacking any satisfactory method of quantifying these matters, all I can do is to record my impressions after long immersion in the period."

*Keith Thomas, The Ends of Life, Oxford University Press, 2009.*

"But the sad truth is that much of what it has taken me a lifetime to build up by painful accumulation can now be achieved by a moderately diligent student in the course of a morning."

*Keith Thomas, Diary, London Review of Books, 10 June 2010.*



Logout

META  **EE Correspondence (critical edition)**

general editor: R. V. McNamee

**Editorial project***Digital correspondence of Voltaire*

general editor: N. Cronk

letter editor: T. D. N. Besterman

published online: 2008

letter n° : D1

**crossref DOI**[https://doi.org/10.13051/ee:  
doc/voltfrVF0850001a1c](https://doi.org/10.13051/ee:doc/voltfrVF0850001a1c)**Writer(s)****Armand Arouet**

writer's age: 19

**Voltaire [François Marie Arouet]**

writer's age: 10

**Recipient(s)****Françoise Aimée Baillif Du Pont**

recipient's age: [unknown]

**Dates**

18 December 1704 (Julian)

29 December 1704 (Gregorian)

Revolutionary not applicable

<< first < previous showing letter 1 of 19604 next > last >> return to [list of letters](#)*Electronic Enlightenment Scholarly Edition of Correspondence*Armand Arouet and Voltaire [François Marie Arouet] to Françoise Aimée Baillif Du Pont<sup>a 1</sup>

Monday, 29 December 1704

DOCUMENT | ENCLOSURES | RELATED | **VERSIONS** | PARENT

Paris, le 29 décembre 1704

Madame et très honorée cousine,

Mon papa m'a fait cette grâce de me comander d'estre son secretaire ce premier d'année, et vous tesmoigner les humbles respects de nostre maison, avec les veux et prières que nous faisons pour vostre prospérité, santé, bonheur et satisfacion, qui ne sont en doutte de vostre costé eu égard à nous. Il vous suplie, madame ma cousine, le croire toujours bon parent et ne vous despartir de l'affecion que vous devez à sa famille, et moy, le secretaire, je finiray en me disant, et Zozo,

Vos très humbles et respectueux cousins,

Zozo  
Arouet<< first < previous showing letter 1 of 19604 next > last >> return to [list of letters](#)

# ***Challenges and opportunities with historical texts and historical language varieties: an example***

Mon papa m'a fait cette grâce de me comander d'estre son secretaire ce premier d'année, et vous tesmoigner les humbles respects de nostre maison, avec les veux et prières que nous faisons pour vostre prospérité, santé, bonheur et satisfacion, qui ne sont en doutte de vostre costé eu égard à nous. Il vous suplie, madame ma cousine, le croire toujours bon parent et ne vous despartir de l'affecion que vous devez à sa famille, et moy, le secretaire, je finiray en me disant, et Zozo,

Vos très humbles et respectueux cousins,

Zozo

Arouet

<https://doi.org/10.13051/ee:doc/voltfrVF0850001a1c>

# *The opportunity...and the problem*

We want to make new forms of research possible by offering freely available online search and exploration of our text collections ...

... but in order to search historical texts effectively, users need to be able to find inflected forms and variant spellings. How can we make that search possible?

<text id="voltfrVF0850001a1c">

<p>

Paris	Paris	NAM	Paris
,	,	PUN	,
le	le	DET:ART	le
29	29	NUM	@card@
décembre	décembre	NOM	décembre
1704	1704	NUM	@card@

</p>

<p>

Madame	Madame	NOM	Madame
et	et	KON	et
très	très	ADV	très
honorée	honorée	ADJ	honoré
cousine	cousine	NOM	cousin
,	,	PUN	,

</p>

<p>

Mon	Mon	DET:POS	mon	
papa	papa	NOM	papa	
m'	m'	PRO:PER	me	
a	a	VER:pres	avoir	
fait	fait	VER:pper	faire	
cette	cette	PRO:DEM	ce	
grâce	grâce	NOM	grâce	
de	de	PRP	de	
me	me	PRO:PER	me	
comander	commander		VER:infi	commander
d'	d'	PRP	de	

estre	être	VER:infi	être
son	son	DET:POS	son
secrétaire	secrétaire	NOM	secrétaire
ce	ce	PRO:DEM	ce
premier	premier	NOM	premier
d'	d'	PRP	de
année	année	NOM	année
,	,	PUN	,
et	et	KON	et
vous	vous	PRO:PER	vous
tesmoigner	témoigner	VER:infi	témoigner
les	les	DET:ART	le
humbles	humbles	ADJ	humble
respects	respects	NOM	respect
de	de	PRP	de
notre	notre	DET:POS	notre
maison	maison	NOM	maison

# Which allows us to search by lemma, pos, original form, or modernized form

Your query "{pouvoir}\_V\*" returned 63,927 matches in 23,424 different texts (in 18,072,242 words [37,655 texts]; frequency: 3,537.30 instances per million words) [1.126 seconds]

< << >> >| Show Page: 1 Line View Show in random order New query Go!

No	Text	Solution 1 to 50	Page 1 / 1279
1	<a href="#">addijoEE0050442a1c</a>	d' hommes , qui est une des plus considerables pertes , que	<a href="#">puisse</a> faire une Colonie naissante , le de
2	<a href="#">addijoEE0050442a1c</a>	en leur envoyant trois ou quatre cents hommes , afin qu' ils	<a href="#">puissent</a> avec ce secours finir la guerre dans
3	<a href="#">addijoEE0060487a1c</a>	Copie de leur accusation , afin que sa defence et sa response	<a href="#">puissent</a> arriver ici et être considerées ense
4	<a href="#">addijoEE0060487a1c</a>	publique prescrite par les dites instructions de Vôtre Majesté , ils ne	<a href="#">pouvoient</a> pas esperer aucun succès . Afin d
5	<a href="#">addijoEE0060514a1c</a>	St . James , et à Madame la Princesse , qu' elle	<a href="#">pouvoit</a> rester dans la Palais , autant qu' el
6	<a href="#">baylpiVF0010001a1c</a>	ne doit avoir aucunement accru votre deplaisir , parce qu' il se	<a href="#">peut</a> faire que lors que l' on leur a écrit
7	<a href="#">baylpiVF0010005a1c</a>	pas aussi celle qu' il faut faire pour si ménager qu' on	<a href="#">puisse</a> être . Ce qui vous a un peu surpris
8	<a href="#">baylpiVF0010005a1c</a>	pour payer l' habit , parce que vous vous attendiés que je	<a href="#">pourrois</a> passer cet été avec l' habit d' hiver
9	<a href="#">baylpiVF0010005a1c</a>	vois que c' est un grand frais , mais pourtant on ne	<a href="#">pût</a> pas s' empêcher en aucune façon
10	<a href="#">baylpiVF0010008a1c</a>	argent que dans les choses necessaires , et dont on ne se	<a href="#">peut</a> pas passer . Vous pourrés ajouter
11	<a href="#">baylpiVF0010008a1c</a>	cheval pour le 6 ou 7 de septembre et afin que je	<a href="#">puisse</a> être au Carla pour le second dima
12	<a href="#">havlniVF0010011a1c</a>	suis depuis long tems . avant énuisé tout le credit que ie	<a href="#">nouveis</a> avoir par 10 ou 12 livres que i' en

# Rhetoric in EE

**Sermocinatio**: the speaker answers the remarks or questions of a pretended interlocutor; or **hypophora**: asking questions and answering them

"Now you might say — Why are you against it ? " and "Were I to say as much Mr White might be in a rage, and Mr Long might say — Mr B. why will you propose such things? — You see it can not be done. The consequence would be, that between Mr White, and Mr Harrison, and Mr Long, possession never would be obtained."

Jeremy Bentham to Sir Evan Nepean, Monday, 9 June 1800

<https://doi.org/10.13051/ee:doc/bentjeOU0060303b1c>

"Je meurs d'envie de savoir de leurs nouvelles et en quels regimens, et sous quels generaux ils ont servi. **Vous me direz que** cela s'en va sans dire, qu'ils ont servi en Hollande dans l'armée meme du Roy puis que le Roy se trouvoit alternativement dans toutes les armées, et que ce qu'on appelloit armée de Mr le prince, de Mr de Turenne, etc. n'etoit proprement parlant que des detachemens de la royale."

Pierre Bayle to Jacob Bayle Monday, 31 July 1673

<http://www.e-enlightenment.com/item/baylpiVF0010218a1c>

## Proverbs and similes

"You must excuse me for giving you a Line of Latin now and then since I find my self in some danger of Losing the Tongue, for I perceive **a new Language, like a New Mistress**, is apt to make a man forget all his old ones."

Joseph Addison to William Congreve, December 1699

<http://www.e-enlightenment.com/item/addijoOU0010010a1c>

**"Mais une femme est un furieux remora; c'est encor pis qu'un prieur"**

Claude Adrien Helvétius to Léger Marie Deschamps Sunday, 7 October 1764

<https://doi.org/10.13051/ee:doc/helvclUT0030145a1c>



# Rhetoric in EE

## Repetition

"Voilà ce que fera votre despote, **ambitieux, prodigue, avare, amoureux, vindicatif, jaloux, foible**: car c' est ainsi qu' ils font tous, et que nous faisons tous. Messieurs, permettez-moi de vous le dire; vous donnez trop de force à vos calculs, et pas assez aux penchans du coeur humain, et au jeu des passions. Votre système est très bon pour les gens de l'Utopie, il ne vaut rien pour les enfans d' Adam."

Jean Jacques Rousseau to Victor Riqueti, marquis de Mirabeau  
Sunday, 26 July 1767 <https://doi.org/10.13051/ee:doc/rousjeVF0330238a>

## Proverbs and similes

"Mais **un royaume est un grand corps isolé**, plus séparé de ses voisins par la diversité d'intérêts, que par la mer, les citadelles, et les barrières qui le renferment."

Beaumarchais to Louis XIV 1717  
<https://doi.org/10.13051/ee:doc/beaupiVF2730014a1cun>

"Tu n'as pas deviné la-dessus, mais bien sur l'humeur de Blaise. **Il étoit comme un dogue ; il est comme un tigre**. Tu as eu bien du pouvoir de faire parler la machine."

Françoise Huguet de Graffigny to François Antoine Devaux  
Thursday, 23 July 1744  
<https://doi.org/10.13051/ee:doc/grafrVF0050722a1c>

**"grasse comme une abbesse"**

Voltaire(twice)

Your query "less \*\_NN2" returned 50 matches in 43 different texts (in 11,422,617 words [1,251 texts]; frequency: 4.38 instances per million words) [0.304 seconds]

< << >> >| Show Page: 1 Line View Show in random order Choose action... Go!

No	Text	Solution 1 to 50	Page 1 / 1
1	<a href="#">S2DD</a> 490 (.) yolo <b>S0688</b> : what can you do ? <b>S0689</b> : what can you do ? <b>S0687</b> : eat	<a href="#">less carbohydrates</a>	and starch I 'm good thank you I 've got a mouthful
2	<a href="#">S2QU</a> 527 <b>S0522</b> : >> who is n't nice <b>S0520</b> : we do n't like him he more or	<a href="#">less rapes</a>	her in the office and then is just a bit self-absorbed and
3	<a href="#">S2XV</a> 425 that they can get on with it (.) but that just engenders more	<a href="#">(.) less students</a>	taking responsibility for their own learning (.) and oh it was just crazy
4	<a href="#">S376</a> 168 thing about next year there 's gon na be less of us	<a href="#">less devices</a>	<b>S0187</b> : yeah still get the same kind of deal or a similar deal
5	<a href="#">S4CU</a> 56 have somehow very cheaply managed to make the engines do this <b>S0521</b> : much	<a href="#">less emissions</a>	<b>S0519</b> : much less emissions how did you do it ? <b>S0521</b> : yeah <b>S0519</b> : erm and
6	<a href="#">S4CU</a> 57 cheaply managed to make the engines do this <b>S0521</b> : much less emissions <b>S0519</b> : much	<a href="#">less emissions</a>	how did you do it ? <b>S0521</b> : yeah <b>S0519</b> : erm and they did n't
7	<a href="#">S57J</a> 420 he had less musical instruments <b>S0510</b> : >> onstage stuff yeah <b>S0509</b> : >> yeah definitely <b>S0510</b> : in fact	<a href="#">less visuals</a>	as well was n't there ? <b>S0509</b> : >> well he did n't have any
8	<a href="#">S5AA</a> 39 n't trust it <b>S0202</b> : yeah <b>S0208</b> : so that 's why there 's been loads	<a href="#">less accidents</a>	because there 's loads less people climbing it <b>S0202</b> : yeah (.) well it says
9	<a href="#">S5SC</a> 340 drunk but <b>S0493</b> : and also you 've got ta remember then there 's	<a href="#">less cubes</a>	in those petri-dishes there to put on the boa- so you can
10	<a href="#">S69V</a> 219 do one is (.) everything you can to reduce the revenue stream <b>S0268</b> : mm	<a href="#">S0269: less turbines</a>	less solar panels <b>S0268</b> : mm <b>S0269</b> : blah blah blah <b>S0268</b> : >> yeah yeah <b>S0269</b> : er the second
11	<a href="#">S6JL</a> 478 you kind of face the fear and the more you go (.) the	<a href="#">less (.) barriers</a>	it seems between you and (.) learning like (.) like u- for me it
12	<a href="#">S784</a> 1310 Hispanic name (.) got ho- erm got ha- erm got (.) way way way	<a href="#">less responses</a>	there is a reason that happens <b>S0402</b> : >> yeah loads of people change their
13	<a href="#">S7JG</a> 1054 prepared (...) <b>S0084</b> : kick it <b>S0041</b> : spit at it (...) <b>S0041</b> : like the women have to do	<a href="#">less press-ups</a>	and stuff do n't they than men ? He 's saying she
14	<a href="#">S8CV</a> 82 we arrived that was what ? five-ish ? <b>S0687</b> : yeah but there were	<a href="#">less cars</a>	than there were there then I think (...) <b>S0688</b> : well <b>S0687</b> : can you sit your
15	<a href="#">S8J6</a> 365 drama <b>S0115</b> : well in some ways that 's good actually (.) cos then with	<a href="#">less dramas</a>	you can focus on you know progression and your work <b>S0037</b> : >> yeah I
16	<a href="#">S8K9</a> 264 're not so many (.) everybody has their own car so there 's	<a href="#">less taxis</a>	(.) unless people drink (.) you know in Tunisia they still drank quite a
17	<a href="#">SB9K</a> 217 <b>S0192</b> : what we have to come up with <b>S0198</b> : yeah <b>S0192</b> : and therefore we write	<a href="#">less words</a>	<b>S0198</b> : I 'm er concerned as well I think --ANONnameM slightly touched on
18	<a href="#">SBDD</a> 140 mean <b>S0115</b> : uh the <b>S0037</b> : so that 's there 's less ma- less and	<a href="#">less matches</a>	are n't there ? Obviously (.) obviously <b>S0115</b> : I think it 's is n't
19	<a href="#">SBG4</a> 318 really ? what about better medicine ? nutrition ? shelter ?	<a href="#">less wars</a>	? less <b>S0417</b> : yeah but I would have thought some Palaeolithic people might
20	<a href="#">SEM7</a> 216 alcohol which would take alcohol money away from the uh <b>S0037</b> : >> yeah (.) exactly	<a href="#">less (.) pubs</a>	and clubs <b>S0115</b> : take alcohol money (.) but you 'd have to look at
21	<a href="#">SH8K</a> 177 there so they get loads of tourists and there 's a lot	<a href="#">less tourists</a>	going so --UNCLEARWORD cos it 's pricier <b>S0422</b> : you do n't understand I
22	<a href="#">SHLT</a> 279 and drops <b>S0024</b> : no (.) yet but the toothbrush nearest to the toilet had	<a href="#">less bacteria</a>	on it than the toothbrush they took outside of the bathroom and
23	<a href="#">SIN6</a> 421 's <b>S0426</b> : the difficulty next year is that we 're gon na have	<a href="#">less members</a>	of staff <b>S0425</b> : mm (.) yeah <b>S0426</b> : and with one of us going on sabbatical
24	<a href="#">SLSS</a> 2476 what that means though do you ? <b>S0584</b> : --UNCLEARWORD <b>S0583</b> : >> look ninety-nine calories or	<a href="#">less Squares</a>	French Fries Whatsits and Quavers (.) then it says a hundred and seven
25	<a href="#">SMW8</a> 4035 marmalade and yes it 's got this and it 's actually got	<a href="#">less ingredients</a>	in than the one she was having so we knew that would





SEARCH

FREQUENCY

CONTEXT

OVERVIEW

(SHUFFLE)

CLICK FOR MORE CONTEXT



[?]

SHOW DUPLICATES

1	1835	NF	EssayOnRateWages	A	B	C	a greater expense?" Yes; though they have their hands for much <b>less wages</b> than in England.' " The average annual wages of labour in England
2	1848	MAG	AmWhigRev	A	B	C	farther. The duty on bread stuffs will be lowered to content the operatives with <b>less wages</b> ; the duty on raw materials for manufacture, to content the owners c
3	1849	MAG	AmWhigRev	A	B	C	our country as cheaply as possible, and if Hessian mercenaries can be had for <b>less wages</b> than free citizens, we should employ them in preference. But here the
4	1852	NF	ElementsPolitical	A	B	C	or diminish, just in proportion as a laborer is able to procure greater or <b>less wages</b> for a day's labor; that is, as every thing is cheaper
5	1862	MAG	NewEngYaleRev	A	B	C	his services for a definite time, works with more indifference and though he receives <b>less wages</b> , he becomes in the end less profitable to his employer. " //
6	1863	FIC	FaithGartneysGirlhood	A	B	C	then we can't afford to pay extravagantly. Mahala has been content to take <b>less wages</b> , and I think she means to be faithful. Perhaps if I make
7	1869	NF	HarrietMosesHer	A	B	C	on the plantation was chosen to go with her. Harriet, who could command <b>less wages</b> than any other child of her age on the plantation, was therefore put
8	1870	MAG	Atlantic	A	B	C	no difficulty but such as patience and firmness can overcome. It will work for <b>less wages</b> than we now pay whites, and its expense for food 478 April,
9	1871	FIC	MillibankRoger	A	B	C	Irving. We will go where you go. We will work for you for <b>less wages</b> than for anybody else, " was what the house servants said to him
10	1876	FIC	WychHazel	A	B	C	' They will make a combination with other mill-owners and undersell him; and paying <b>less wages</b> they can afford to do it, for a time. And a certain
11	1879	NEWS	NYT-Reg	A	B	C	Blaine. Because the Chinese could eat rice and rats, and could work for <b>less wages</b> than the American laborer, was not sufficient cause for abrogating the treaty
12	1887	MAG	NewEngYaleRev	A	B	C	majority of the great capitalists of to-day were twenty and twenty-five years ago working for <b>less wages</b> per diem than the common day laborer is earning to-da
13	1890	MAG	NorthAmRev	A	B	C	either have more markets and more consumers, or less product, less employment, <b>less wages</b> , and less profits to capital. Situated as we are to-day, we
14	1892	NF	WhiteSlavesOppression	A	B	C	. The incoming of the sweater has not resulted in more work, but in <b>less wages</b> . If my critic will take the trouble to examine the testimony given before
15	1892	NF	WhiteSlavesOppression	A	B	C	side by side with men in the same shops and stores, are paid far <b>less wages</b> for the same work. This is an aristocracy of sex that shames and
16	1894	MAG	NorthAmRev	A	B	C	that it showed the needs of labor and enabled the employers to obtain men for <b>less wages</b> than would otherwise be paid. The same complaint, it may be notec
17	1895	FIC	JCole	A	B	C	" who was a " verry good hite, " and the offer to take <b>less wages</b> if " I would do his washin, " I found myself wondering what
18	1896	FIC	FameFortune	A	B	C	of the most prosperous merchants in the city began in that way, only on <b>less wages</b> . One who wants to climb the ladder of success must, except in
19	1906	FIC	RandyRiver	A	B	C	won't try, so I have the work, an' the girrul gits <b>less wages</b> , " answered the Irishwoman. When Randy returned home he found supper almost
20	1909	FIC	WhatDianthaDid	A	B	C	am willing to undertake these intricate and responsible duties for a seven day week at <b>less wages</b> than are given the street-digger, for \$1.50 a day. " " Good
21	1912	MAG	Atlantic	A	B	C	some of them have already gone, and where they now work fifty-seven hours for <b>less wages</b> than the Lawrence mills pay for fifty-four hours? This is not meant
22	1913	MAG	Atlantic	A	B	C	as is reported, against the white unions, so that they may pay them <b>less wages</b> , it is likewise true that some white unions take advantage of the existing
23	1913	NEWS	NYT-Reg	A	B	C	conductors. The railroads asked that extra men made necessary by full crew laws get <b>less wages</b> , but the board says that it is " not its function to correct
24	1922	NEWS	WallStjrnl	A	B	C	the teeth of the competition of other manufacturing nations she must work longer hours for <b>less wages</b> , she must cut profits, she must reduce her imports to t
25	1925	NF	LeadingConstitutional	A	B	C	in the particular occupations under normal conditions. Provision is made for special licenses at <b>less wages</b> in the case of women who are incapable of full servi
26	1958	NEWS	WallStjrnl	A	B	C	national product (total value of goods and services produced by the entire economy, <b>less wages</b> paid to all government employes) in terms of 1947 dollars divid



Repository

Corpus Search

TreeQuery

Treex

More Apps

About



Login



Query Corpora Save Concordance Filter Frequency Collocations View Help

Corpus: EU DGT-UD: English | Query: less (66,677 hits) | Shuffle: ✓

1

### Text types overview

### Usage tips

In case your CQL query produces nested matches (e.g. if you search one or more adjectives preceding a noun) you can use 'Filter -> Remove nested matches'.

Hits: 66,677 | i.p.m.: 440.63 (related to the whole corpus) | ARF: 21,577.16 | Result is shuffled

1 / 1,667

Line selection: simple




<input type="checkbox"/>	32015R1754 + 2015	€ 107,1 Not less than € 102,8 but	less	than € 105 Not less than € 100,7
<input type="checkbox"/>	32016R1969 + 2016	± 100 Hz of a final specified output frequency of	less	than 1 GHz ; or b. ± 0,1 part per
<input type="checkbox"/>	32017R0543 + 2016	1,5 – less than 2,0 rooms per occupant 2,0 –	less	than 2,5 rooms per occupant 2,5 – less than
<input type="checkbox"/>	32008R0008 + 2008	with a maximum approved passenger seating configuration of 9 or	less	, and a maximum take - off mass of 5700
<input type="checkbox"/>	32008R1031 + 2008	€ 84,6 Not less than € 81,2 but	less	than € 82,9 Not less than € 79,5 but
<input type="checkbox"/>	42012X0508(01) + 2012	100 mm diameter shall be rounded with a radius not	less	than 2,5 mm ; those protruding less than 5 mm
<input type="checkbox"/>	32004R1810 + 2004	or more ( not exceeding 12 metric number ) Measuring	less	than 833,3 decitex but not less than 277,8 decitex
<input type="checkbox"/>	32006D0599 + 2006	the borrower can obtain lower rates and / or offer	less	security . In some cases , the borrower would not
<input type="checkbox"/>	42011X0909(01) + 2011	in the test referred to in paragraph 6.3.2 and not	less	than 1470 daN . The Technical Service conducting the te
<input type="checkbox"/>	31992L0116 + -2003	: 1 litre per carcass weighing 2,5 kg or	less	, 1,5 litres per carcass weighing between 2,5 kg
<input type="checkbox"/>	32008R1031 + 2008	per litre With an octane number ( RON ) of	less	than 98 Spirit type jet fuel Other light oils Medium
<input type="checkbox"/>	31988R0222 + -2003	of a free - at - frontier value of not	less	than 365,76 ( 2 ) Glarus herb cheese ( known as
<input type="checkbox"/>	32005R1719 + 2005	vanadium , more than 1,2 % of carbon and not	less	than 11 % but not more than 15 % of
<input type="checkbox"/>	32004R1810 + 2004	fermentation of wine and having a total acidity of not	less	than 60 g/l , expressed as acetic acid . Waters
<input type="checkbox"/>	32012R0388 + 2012	" system angular beam steering errors of equal to or	less	than 10 µrad ; Optical equipment and components , spec
<input type="checkbox"/>	32005L0078 + 2005	process of an emission control device that occurs periodically in	less	than 100 hours of normal engine operation . During cycle
<input type="checkbox"/>	32016R1821 + 2016	€ 66,2 but less than € 67,6 Not	less	than € 64,9 but less than € 66,2 Not
<input type="checkbox"/>	32015R2420 + 2015	time resolution ; Framing cameras capable of 50 ns or	less	frame exposure time ; 3 . Framing tubes and solid
<input type="checkbox"/>	32007R1184 + 2007	the Community consumption ( less than 4 % ) and	less	than 8 % of the sales of the Community producer
<input type="checkbox"/>	31994L0055 + -2003	organic peroxide and which have a boiling point of not	less	than 150 °C . Type A diluents may be used
<input type="checkbox"/>	32013R1001 + 2013	+ 4,6 € /100 kg /net [ 10 ] Not	less	than € 40,7 but less than € 41
<input type="checkbox"/>	32007R1183 + 2007	. scatterometers having a measurement accuracy of 10 ppm or	less	( better ) ; profilometers having a measurement accuracy



11 / 206 korpusta valittuina – 161,94M / 3,45G sanetta

less 

Yksinkertainen Laajennettu Edistynyt Vertailu

less Hae myös  alkuosa  loppuosaa ja  samaista pien- ja suuraakkosetvirkkeistä  jotka sisältävät Konkordanssi: osumia sivulla: 25  järjestä korpuksen sisällä: järjestämätön  Tilastoja: laskentaperuste: sana   Näytä kartta

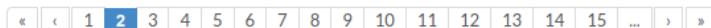


Konkordanssi Tilastoja Kartta

Tuloksia: 38 213

 Näytä konteksti

AGRICULTURE AND FORESTRY (ei tue laajennettua kontekstia)

Basically, uropean consumers know what elements belong to a health-promoting diet, " more fruit and vegetables", " less fat", " balance and va iety' ( Margetts et al. However, a recent meta-analysis indicates a **less** According to a Belgian study, consumers are even **less** willing to compromise on the taste of functional foods than previously ( Verbeke 2006 ). ent was familia, the increasing strength of the claim did not increase the perceived benefit but if the component was **less** familiar, the percei ed advantage improved with the increasing strength of the The soup with  $\beta$ -glucan had a **less** intense orange colour, the freshness of the aroma, the intensity of the shrimp aroma, the shri p flavour, the d than  $\beta$ -glucan, in the preparations may also have an effect on the texture characteristics but probably their effects ar **less** than the effect of molecular weight or cereal source of  $\beta$ -glucan. ther study, French consumers have been found to seek more pleasure fro food and to prefer low-fat or low-salt diets **less** than Belgians, Americans or Japanese ( Rozin et al. In a Danish study, American and Danish consumers were **less** positive than Finnish consumers about the foods enriched with omega-3 fatty acids or oligosaccharides, but In another Danish study, Danish consumers knew **less** about functional foods and had more doubts about their benefits th n English consumers ( Jonas and Beckm: Business opportunities that are closely related to existing capabilities involve **less** risk, but they are also unlikely to change the current performance ( Hoskisson and Busenitz 2001 ). The **less** complex the organisation, the easier it is to monitor communication, increase costeffectiveness, and in some Further, it has been found that strengthening centralisation and formalisation implies **less** involvement ( personal a countability ), which in itself is a source of intrinsic motivation and a key driver of er **less** planning flexibility, which would decrease the reliability of routines, and liquidity and, I believe, no **less** useful " ( Richardson 1972: 888 ) According to Nelson and Winter ( 1982 ), skills, organisationand ' echnology ot limited to milking robots, but the need to p rform automated monitoring is the greatest in automated systems with **less** human contact ( Maltz & Spahr 1997 ). They found that discomfort unde a hoof caused cows to keep **less** weight on the affected leg and shift weight to the contra lateral leg. The variability of the weight between legs may be a better indicator than just a leg bearing **less** eight ( Neveux et al. It was found that the output of the sensor is linear and the location of the oad had **less** than a 1 kg effect on the output of the sensor. If **less** than 90 s of the original weighing remained after removing erroneous data, the whole weighing was discard A " kick " was described as lifting of the leg and calculated when the weight of a leg decreased to **less** than 5 kg in order to avoid the effect of possible drift in the zero level ) and a " step " was described as weight The figures show that in the mild stage ( on 10/21, Figure 10 ) the cow had started to place **less** weight on the affecte leg ( Figure 11b ). The fact that the behaviour shown in the cows in this study is similar to that reported in cows by others, e.g. **less** weigh on the affected leg ( Neveux et al. 5 ), who have also utilised a system measuring leg weights and found that discomfort under hoof causes a cow to bear **less** weight on the affected leg. cumu ated in TSA plategrown cultures ( B. cereus emetic strains NS117 and 5964a ) when the atmosphere contained **less** than 1.6 vol % O<sub>2</sub>, but when the O<sub>2</sub> concentration was 4.5 vol %, high amounts ( about 1000 ng mg-1 ) of cer The method thus required solvents highly toxic to humans ( tetrahydrofuran, chloroform ), whereas we used **less** hazardous solvents.

Lataa konkordanssi muodossa Annotaatiot  Excel (XLS) **Korpus**

Agriculture and Forestry

Viittaa aineistoon

**Tekstin piirteet**

tekstin otsikko: Antimicrobial Compounds and Extracellular Polysaccharides Produced by Lactic Acid Bacteria: Their Structures and Properties  
päivämäärä: 2000-03  
avainsanat: [tyhjää]  
tiedekunta: University of Helsinki, Faculty of Agriculture and Forestry, Department of Food Technology  
oppiaine: [tyhjää]  
tutkielman tyyppi: Doctoral dissertation

Lataa työ E-thesiksestä



# WORD SKETCH

English Web 2015 (enTenTen15)

Get more space [+](#) [↗](#) [?](#) [!](#) [⋮](#)

less as adjective 2,466,667× ...

[🔍](#) [⬇️](#) [👁️](#) [🌐](#) [ⓘ](#) [☆](#)

modifiers of "less"
<b>no</b> ...
no less than
<b>nothing</b> ...
is nothing less than a
<b>considerably</b> ...
considerably less than
<b>much</b> ...
much less
<b>significantly</b> ...
significantly less than
<b>far</b> ...
far less than
<b>little</b> ...
a little less than
<b>slightly</b> ...
slightly less than
<b>substantially</b> ...
substantially less than
<b>somewhat</b> ...
somewhat less than
<b>even</b> ...
even less
<b>preferably</b> ...

nouns modified by "less"
<b>money</b> ...
less money
<b>energy</b> ...
less energy
<b>attention</b> ...
less attention
<b>stress</b> ...
less stress
<b>time</b> ...
less time
<b>water</b> ...
less water
<b>chance</b> ...
less chance of
<b>waste</b> ...
less waste
<b>fuel</b> ...
less fuel
<b>electricity</b> ...
less electricity than
<b>pain</b> ...
less pain
<b>risk</b> ...

"less" and/or ...
<b>less</b> ...
less and less
<b>more</b> ...
more or less
<b>great</b> ...
greater or less than
<b>cheap</b> ...
the cheaper and less
<b>slow</b> ...
slower and less
<b>easy</b> ...
easier and less time consuming
<b>small</b> ...
smaller and less
<b>fast</b> ...
a faster and less
<b>weak</b> ...
weaker and less
<b>short</b> ...
shorter and less
<b>simple</b> ...
a simpler and less
<b>old</b> ...

prepositional phrases
"less" than ...
"less" of ...
"less" in ...
"less" on ...
"less" about ...
"less" for ...
"less" per ...
"less" at ...
"less" like ...
"less" from ...
"less" to ...
"less" with ...

infinitive objects of "less"
<b>fear</b> ...
less to fear from
<b>worry</b> ...
less to worry about
<b>lose</b> ...
less to lose
<b>taste</b> ...
more or less to taste
<b>operate</b> ...
cost less to operate
<b>spend</b> ...
less to spend on
<b>heat</b> ...
less to heat and
<b>do</b> ...
less to do with
<b>manufacture</b> ...
cost less to manufacture
<b>complain</b> ...
have less to complain about
<b>cast</b> ...
less to cast
<b>fuel</b> ...

verbs complemented by "less"
<b>expect</b> ...
expect nothing less
<b>cost</b> ...
cost you less
<b>accept</b> ...
accept anything less than
<b>pay</b> ...
<b>take</b> ...
<b>give</b> ...
<b>deserve</b> ...
deserve nothing less than
<b>have</b> ...
<b>love</b> ...
<b>charge</b> ...
<b>need</b> ...
need nothing less than
<b>receive</b> ...

verbs before "less"
<b>have</b> ...
have less than
<b>cost</b> ...
cost less than
<b>take</b> ...
takes less than
<b>earn</b> ...
earn less than
<b>pay</b> ...
pay less
<b>spend</b> ...
spend less
<b>care</b> ...
could n't care less
<b>weigh</b> ...
weighs less than
<b>receive</b> ...
received less than
<b>last</b> ...
lasted less than
<b>use</b> ...
use less
<b>repr</b> ...

[Back to the original interface](#)

# D\*/zeit: DiaCollo



QUERY: Maus

DATE(S):  SLICE:

SCORE: log Dice (ld)  KBEST:  CUTOFF:

PROFILE: collocations  FORMAT: Cloud  GLOBAL:

GROUPBY:  DEBUG:

[Home](#) [Info](#) [Help](#)

1.000x



# Corpus analysis functions

- Searching
  - words
  - lemmas
  - wildcards
  - regular expressions
  - annotations
- Concordancing
- Collocations
- Expanding co-text
- Displaying annotations
- Using metadata
- Wordlists
- Keywords
- Identifying multi-word units
- Parallel text alignment
- Beyond text: audio, video and images...

<https://ota.ox.ac.uk/documents/searching/handbook.html>

# Annotation

Annotation of texts should include structural markup, metadata, and linguistic annotation, including:

- Standardized metadata for basic categories such as language, relevant dates, author, title and text type;
- Part-of-speech tagging;
- Lemmatization; and
- Modernized (or otherwise normalized) forms

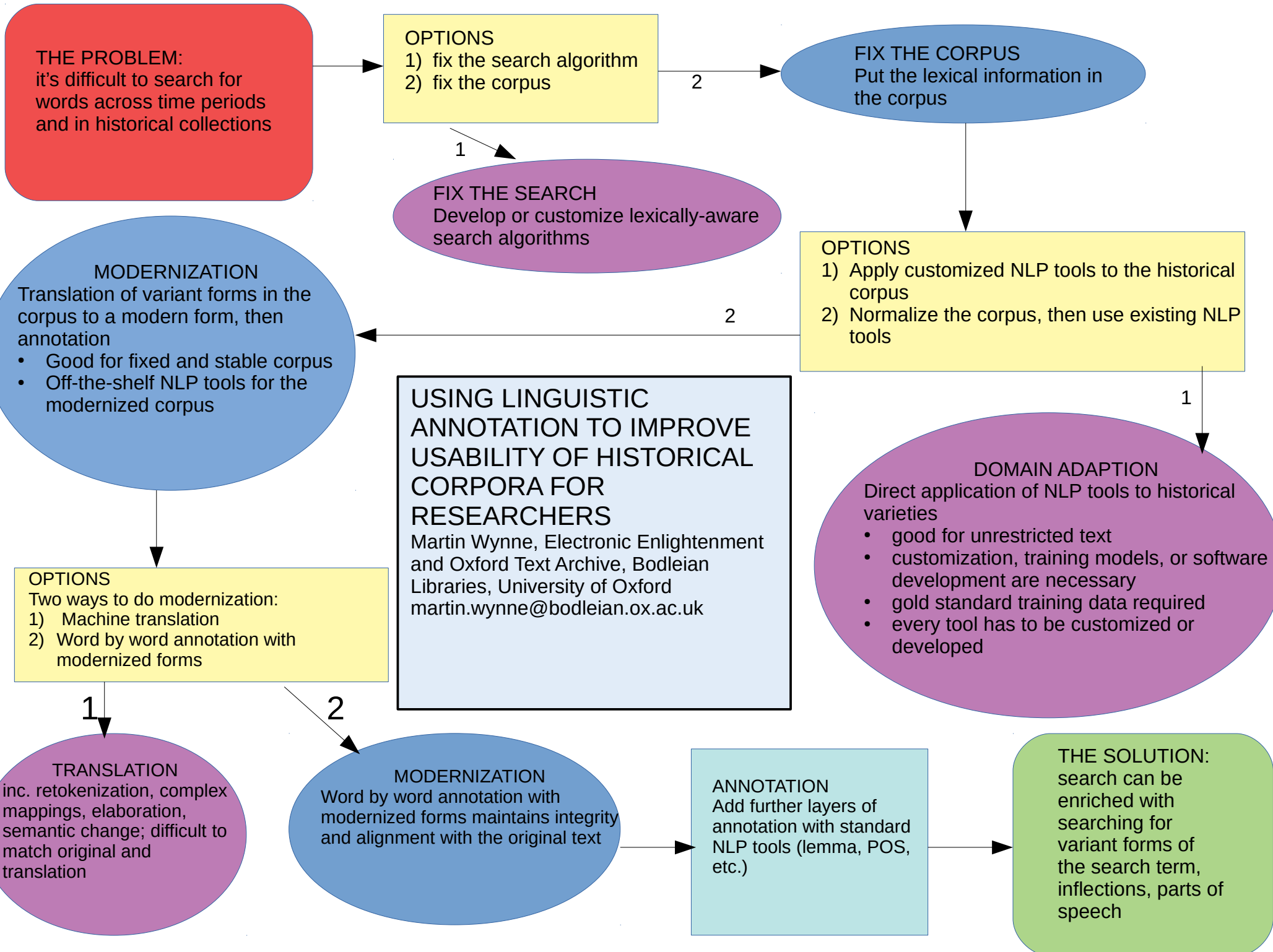
...and these can be the basis for further levels of annotation, such as:

- semantic tags
- named entity recognition
- etc.

# Annotation and corpus interfaces make possible...

- **Enhanced search**, for inflected and variant forms of words;
- **Cross-searching** the content of numerous texts, with results refined on the basis of content and metadata values;
- **Visualization** of results;
- **'Distant reading'**, data mining and extraction of results across large numbers of texts;
- **Limited access** via search and exploration interfaces to **restricted texts**, which would otherwise not be available to readers;
- Application of **further levels of annotation** and analysis by users based on the basic levels of annotation;
- Linking to resources held in other repositories, allowing cross-repository resource discovery and content search – see CLARIN **Federated Content Search** prototype (<https://contentsearch.clarin.eu/>)





**THE PROBLEM:**  
it's difficult to search for words across time periods and in historical collections

**OPTIONS**  
1) fix the search algorithm  
2) fix the corpus

**FIX THE CORPUS**  
Put the lexical information in the corpus

**FIX THE SEARCH**  
Develop or customize lexically-aware search algorithms

**OPTIONS**  
1) Apply customized NLP tools to the historical corpus  
2) Normalize the corpus, then use existing NLP tools

**MODERNIZATION**  
Translation of variant forms in the corpus to a modern form, then annotation  
• Good for fixed and stable corpus  
• Off-the-shelf NLP tools for the modernized corpus

**USING LINGUISTIC ANNOTATION TO IMPROVE USABILITY OF HISTORICAL CORPORA FOR RESEARCHERS**

Martin Wynne, Electronic Enlightenment and Oxford Text Archive, Bodleian Libraries, University of Oxford  
martin.wynne@bodleian.ox.ac.uk

**DOMAIN ADAPTION**  
Direct application of NLP tools to historical varieties  
• good for unrestricted text  
• customization, training models, or software development are necessary  
• gold standard training data required  
• every tool has to be customized or developed

**OPTIONS**  
Two ways to do modernization:  
1) Machine translation  
2) Word by word annotation with modernized forms

**TRANSLATION**  
inc. retokenization, complex mappings, elaboration, semantic change; difficult to match original and translation

**MODERNIZATION**  
Word by word annotation with modernized forms maintains integrity and alignment with the original text

**ANNOTATION**  
Add further layers of annotation with standard NLP tools (lemma, POS, etc.)

**THE SOLUTION:**  
search can be enriched with searching for variant forms of the search term, inflections, parts of speech

# Links

<http://ota.ox.ac.uk/>

<http://www.e-enlightenment.com/>

<http://digital.bodleian.ox.ac.uk/>

<http://www.clarin.eu/>

<https://cqpweb.lancs.ac.uk/>

<http://kaskade.dwds.de/dstar/dta/diacollo/>

<https://www.sketchengine.eu/>

<https://lindat.mff.cuni.cz/services/kontext/corpora/corplist>

<https://www.english-corpora.org/>