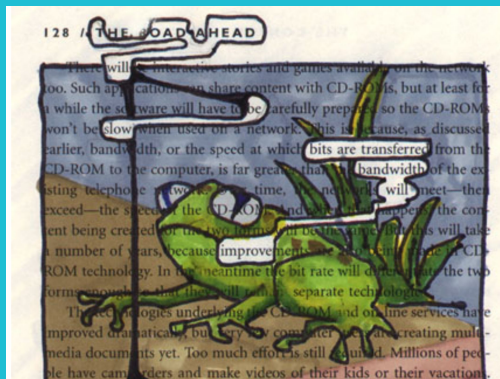


/instituut voor
de Nederlandse
taal/

Out of the Frying-Pan into the Fire

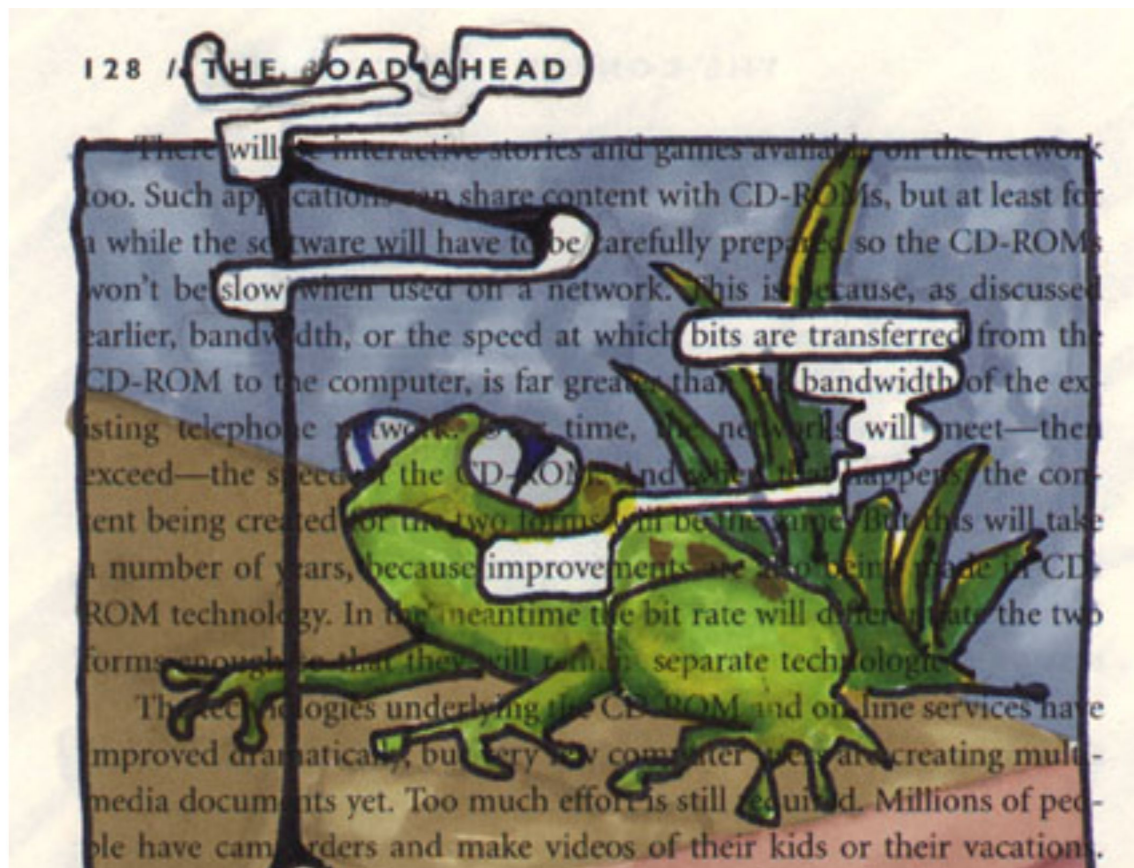
corpus annotation



Katrien Depuydt, Jesse de Does
katrien.depuydt@ivdnt.org
jesse.dedoes@ivdnt.org

Out of the Frying-Pan into the Fire

corpus annotation



Corpus annotation layers

- *Basic layers:*
 - lemma
 - part of speech
 - named entities
 - recognition
 - linking
- *Syntax:* treebanks
- *Semantics/pragmatics:*
 - word sense annotation
 - coreference

Annotation tools

- PoS taggers (+ lemmatizers), NER, parsing, semantic tagging, ...
 - Frog <https://languagemachines.github.io/frog/> (online version in TTNWW)
 - OpenConvert <http://openconvert.clarin.inl.nl> (modern Dutch)
 - “INL labs” <http://inl-labs.inl.nl> (historical dutch)
- Run on own hardware / Web-service based

Input and output

- Input: usually plain text or some XML
 - TEI <http://www.tei-c.org>
 - FoLiA <https://proycon.github.io/folia/>
- Output: (verbose) XML or plain text

1	Marie	Marie	[Marie]	SPEC(deeleigen)	1.000000	B-PER	B-NP	2	su
2	vroeg	vragen	[vraag]	WW(pv,verl,ev)	0.532544	0	B-VP	0	ROOT
3	zich	zich	[zich]	VNW(refl,pron,obl,red,3,getal)	0.999740	0	B-NP	2	se
4	af	af	[af]	VZ(fin)	0.996853	0	0	2	svp
5	of	of	[of]	VG(onder)	0.733333	0	B-SBAR	4	vc
6	hij	hij	[hij]	VNW(pers,pron,nomin,vol,3,ev,masc)	0.999659	0	B-NP	8	su
7	nog	nog	[nog]	BW()	0.999930	0	B-ADVP	8	mod
8	zou	zullen	[zal]	WW(pv,verl,ev)	0.999947	0	B-VP	5	body
9	komen	komen	[kom][en]	WW(Inf,vrij,zonder)	0.861549	0	I-VP	8	vc
10	.	.	[.]	LET()	0.999956	0	0	9	punct

example PoS tagging: modern

```

<s>
<w lemma="de" type="PD(type=d-p,subtype=art-def)" xml:id="w.0">De</w>
<w lemma="tovenaar" type="NOU-C(gender=f|m,number=sg)" xml:id="w.1">tovenaar</w>
<w lemma="en" type="CONJ(type=coor)" xml:id="w.2">en</w>
<w lemma="de" type="PD(type=d-p,subtype=art-def)" xml:id="w.3">de</w>
<w lemma="hobbit" type="NOU-C(gender=f|m,number=sg)" xml:id="w.4">hobbit</w>
<w lemma="duwen" type="VRB(finiteness=fin,mood=ind,tense=past,number=pl)" xml:id="w.5">duwden</w>
<w lemma="het" type="PD(type=d-p,subtype=art-def)" xml:id="w.6">het</w>
<w lemma="zwaar" type="AA(degree=pos,position=prenom,formal=infl-e)" xml:id="w.7">zware</w>
<w lemma="krakend" type="AA(degree=pos,position=prenom,formal=infl-e)" xml:id="w.8">krakende</w>
<w lemma="hek" type="NOU-C(gender=n,number=sg)" xml:id="w.9">hek</w>
<w lemma="open" type="AA(degree=pos,position=adv|pred)" xml:id="w.10">open</w>
<w lemma="en" type="CONJ(type=coor)" xml:id="w.11">en</w>
<w lemma="lopen" type="VRB(finiteness=fin,mood=ind,tense=past,number=pl)" xml:id="w.12">lopen</w>
<w lemma="over" type="ADP(type=pre)" xml:id="w.13">over</w>
<w lemma="een" type="PD(type=indef,subtype=art-indef)" xml:id="w.14">een</w>
<w lemma="breed" type="AA(degree=pos,position=prenom)" xml:id="w.15">breed</w>
<w lemma="pad" type="NOU-C(gender=n,number=sg)" xml:id="w.16">pad</w>
<w lemma="naar" type="ADP(type=pre)" xml:id="w.17">naar</w>
<w lemma="het" type="PD(type=d-p,subtype=art-def)" xml:id="w.18">het</w>
<w lemma="huis" type="NOU-C(gender=n,number=sg)" xml:id="w.19">huis</w>
</s>
<pc type="post" xml:id="pc.000001">.</pc>

```

De tovenaar en de hobbit duwden het zware krakende hek open en liepen over een breed pad.

word	pos	lemma
De	PD(type=d-p,subtype=art-def)	de
tovenaar	NOU-C(gender=f m,number=sg)	tovenaar
en	CONJ(type=coor)	en
de	PD(type=d-p,subtype=art-def)	de
hobbit	NOU-C(gender=f m,number=sg)	hobbit
duwden	VRB(finiteness=fin,mood=ind,tense=past,number=pl)	duwen
het	PD(type=d-p,subtype=art-def)	het
zware	AA(degree=pos,position=prenom,formal=infl-e)	zwaar
krakende	AA(degree=pos,position=prenom,formal=infl-e)	krakend
hek	NOU-C(gender=n,number=sg)	hek
open	AA(degree=pos,position=adv pred)	open
en	CONJ(type=coor)	en
liepen	VRB(finiteness=fin,mood=ind,tense=past,number=pl)	lopen
over	ADP(type=pre)	over

Part of speech tagging: historical



INL labs α

Help

Home

Instituut voor Nederlandse Lexicologie

title:

DE NEDERLANDSE TAAAL

¶t waeter, doch leevendigh daaruit; hoewel zulx bezeert van den slaghen en de splinters, dat nocht hy, nocht zyn vendrigh weder t'scheepe kon koomen. ¶Romero, met een deel anderen, haastte zich oover wegh, ongeduldigh in de traagheit van Glimes; den welken vaarende met bedachtzaamer zinnen, hy echter moest inwachten, en zetten 't voor Roemerswaal. ¶Louis Boisot, Ammiraal van Zeelandt, doende 's vyands beweeghenis vlytelyk waarneemen, en hier af verkundschapt, zond daatlyk de Hopluiden Joost de Moor, Jacob Stoffelszoon, en Vincent Neuz, met acht van de lichtste scheepen, ter schermutzinge. ¶Midlerwyl hield hy kryssaadt, en werd beslooten, dat men, hebbende reeds 't voordeel van een Noordwesten windt, teeghens d'achtere op zouw streeven, om' met een' voorvloedt, aan den man te koomen; zonder langer vertrek: te meer, naardien vier van de weerbaarste scheepen door den Prins te rug ontbooden waaren, en neevens andere veerigh van den onderaanstaen.

van	ADP	van	(van:NOU, van:CON, vinden:VRB)
den	ART	de	(den:NOU)
slaghen	NOU	slag	
en	CON	en	(en:ADV, en:ADP, zij:PRN, en:?)
de	ART	de	(doen:VRB)
splinters	NOU	splinter	
dat	CON	dat	(dat:CON, dat:NOU)
nocht	CON	nocht	
hy	PRN	hij	(hei:INT, hij:?)
nocht	CON	nocht	
zyn	VRB	zijn	(zijn:PRN, zijn:NOU, zijne:NOU)
vendrigh	NOU	vendrigh	
weder	ADV	weder	(weder:NOU, weer:NOU)
t'	ART	het	(het:ART)

Trefwoorden voor: vendrig [1]

Periode
1500-heden [1]

1500-heden

• vendrig ("vendrig") zelfstandig naamwoord (mannelijk)

1. Oorspr. bij de infanterie: jong officier, meestal van aanzienlijke geboorte, die het vaandel van de compagnie droeg. Naderhand: jongste officier der compagnie. [↗](#)
2. (Barg.) Soldaat. [↗](#)
3. (Zndl.) In niet-militair gebruik: drager van het vaandel. [↗](#)

Example text

en den heer Directeur
van het Bederlaers Werkhuis
Brugge.

Bruggen den 27 Maert 1864

Mynheer

Den ondergeteekende op gesloten in het
bedelaers-werkhuis van Brugge sedert den
28 Oogst 1862 heeft de eer
UEu edele te verzoeken
my te willen voorstellen aen den heer
Burgmeester van Bruggen voor het bekomen
myner vryheid. *Zynde nu juist in het goed
voor hetgeen mynen stiel betreft
ik durf my op uwe gewoone
welwillend heid betrouwen om eene goede
antwoord op mynen brief te bekomen
In afwachting blyve ik met den
grootsten eerbied*

Uwen onderdanigen Dienaer
Franciscus Hendrickx

Zynde zynde N(soort,ev,basis,zijd,stan)
nu nu BW()
juist juist BW()
in in VZ(init)
het het LID(bep,stan,evon)
goed goed ADJ(vrij,basis,zonder)
voor voor VZ(init)
hetgeen hetgeen VNW(betr,det,stan,nom,zonder,zonder-n)
mynen mynen N(soort,mv,basis)
stiel stiel N(soort,ev,basis,zijd,stan)
betreft betreffen WW(pv,tgw,met-t)
ik ik VNW(pers,pron,nomin,vol,1,ev)
durf durven WW(pv,tgw,ev)
my my SPEC(vreemd)
op op VZ(init)
uwe uw VNW(bez,det,dial)
gewoone gewoone N(soort,ev,basis,zijd,stan)
welwillend welwillend ADJ(vrij,basis,zonder)
heid heid WW(od,vrij,zonder)
betrouwen betrouwen WW(inf,vrij,zonder)
om om VZ(init)
eene eene ADJ(prenom,basis,met-e,stan)
goede goed ADJ(prenom,basis,met-e,stan)
antwoord antwoord N(soort,ev,basis,onz,stan)
op op VZ(init)
mynen mynen N(soort,mv,basis)
brief brief N(soort,ev,basis,zijd,stan)
te te VZ(init)
bekomen bekomen WW(inf,vrij,zonder)
In in VZ(init)
afwachting afwachting N(soort,ev,basis,zijd,stan)
blyve blyve WW(pv,verl,ev)
ik ik VNW(pers,pron,nomin,vol,1,ev)
met met VZ(init)
den de LID(bep,dial)
grootsten grootsten ADJ(dial)
eerbied eerbied N(soort,ev,basis,zijd,stan)

Zynde VRB zijn
nu ADV nu
juist ADV juist
in ADP in
het ART het
goed NOU goed
voor ADP voor
hetgeen PRN hetgeen
mynen PRN mijn
stiel NOU stiel
betreft VRB betreffen
ik PRN ik
durf VRB durven
my PRN ik
op ADP op
uwe PRN uw
gewoone ADJ gewoon
welwillend NOU welwillend
heid NOU heide
betrouwen VRB betrouwen
om CON om
eene ART een
goede ADJ goed
antwoord NOU antwoord
op ADP op
mynen PRN mijn
brief NOU brief
te ADP te
bekomen VRB bekomen
In ADP in
afwachting NOU afwachting
blyve VRB blijven
ik PRN ik
met ADP met
den ART de
grootsten ADJ groot
eerbied NOU eerbied

Tagger invocation example

```
import scala.xml._

object Tagger
{
  def bab_tagger_url(text:String):String =
    s"http://inl-labs.inl.nl/succeed/text?input=${java.net.URLEncoder.encode(text)}&format=text&tagger=bab-tagger&output=raw"
  def chn_tagger_url(text:String):String =
    s"http://openconvert.clarin.inl.nl/openconvert/text?input=${java.net.URLEncoder.encode(text)}&format=text&tagger=chn-tagger&output=raw&to=tei"

  def getTextButNotIn(x:Node,tagName:String):String =
    if (x.isInstanceOf[Text])
      x.text
    else
      x.child.filter(_label != tagName).map(getTextButNotIn(_,tagName)).mkString("")

  def tag(f:String=>String, text:String) =
    {
      val d = XML.load(f(text))
      val z = (d \ "w").map(w =>
        List(
          ("word", getTextButNotIn(w,"interp")),
          ("lemma", (w \ "@lemma").text),
          ("pos", (w \ "@type").text)))
      z.foreach(println)
    }
}
```

```
Tagger.tag(Tagger.bab_tagger_url,"De Aerde nu was woest ende ledich, ende duysternisse was op den afgront: ende de Geest Godts sweefde op de Wateren.")
```



Named entity recognition: “modern”

Die dag zongen ze niet en vertelden ook geen verhalen, hoewel het weer beter werd; en ook de volgende dag en de dag daarna niet. Ze voelden nu dat het gevaar aan weerskanten niet ver weg was. Zij overnachtten onder de sterren en hun paarden kregen meer te eten dan zij, want er was volop gras, maar er zat niet veel in hun zakken, ook niet met dat wat zij van de trollen hadden genomen. Op een morgen doorwaadden zij de rivier op een brede, ondiepe plek vervuld van het lawaai van ratelende stenen en brui - send schuim. De andere oever was steil en glibberig. Toen zij de top ervan hadden bereikt, hun pony's leidend, zagen zij dat zij de grote bergen heel dicht genaderd waren. Zij schenen nu nog slechts een dag gaans van de voet van de dichtstbijzijnde verwijderd te zijn. Hij zag er donker en naargeestig uit, hoewel er vlekken zonlicht op de bruine hellingen lagen, en achter zijn rondingen glinsterden de toppen van sneeuwkapen. 'Is dat **De Berg**?' vroeg **Bilbo** op plechtige toon, terwijl hij er met ronde ogen naar keek. Hij had nog nooit eerder zo iets groots gezien.

'Natuurlijk niet!' zei **Balin**. Mat is slechts het begin van de **Nevelbergen** en wij moeten er op de een of andere manier door, of erover, of eronder voor - dat wij **Wilderland** erachter kunnen bereiken. En het is ook van de andere kant nog een behoorlijk eind naar de **Eenzame Berg** in het oosten waar **Smaug** op onze schat ligt.

'O!' zei **Bilbo** en op dat ogenblik voelde hij zich vermoeider dan hij zich ooit had gevoeld. Hij moest weer denken aan zijn luie stoel voor de haard in de geliefde zitkamer van zijn hobbit-hol en aan de ketel die zong. En niet voor de laatste keer!

Nu ging **Gandalf** voorop. 'Wij mogen de weg niet mislopen, anders zijn we er geweest,' zei hij. 'Wij hebben in de eerste plaats eten nodig, en rust in re - delijke veiligheid - ook is het absoluut noodzakelijk om de **Nevelbergen** langs het goede pad in te trekken, anders raak je erin verdwaald en moet je teruggaan en weer helemaal opnieuw beginnen (als je tenminste nog te - rugkomt).' Ze vroegen hem waar hij heenging en hij antwoordde: 'Jullie zijn nu vlak bij de rand van de Wildernis gekomen, zoals sommigen van jullie wellicht weten. Ergens voor ons, verscholen, ligt het schone dal Ri - vendell, waar **Elrond** woont in het Laatste Huiselijke Huis. Ik heb mijn vrienden een boodschap meegegeven en wij worden verwacht.'

Namescape: NER for literary studies

tagger	NE type	precision	recall	F1
Stanford	location	0.802	0.712	0.754
	misc	0	0	0
	organisation	0.433	0.228	0.299
	person	0.876	0.895	0.881
	overall	0.853	0.824	0.838
Namescape	location	0.83	0.729	0.776
	misc	0	0	0
	organisation	0.516	0.251	0.339
	person	0.867	0.917	0.896
	overall	0.853	0.838	0.845
Namescape (+wv)	location	0.858	0.830	0.844
	misc	0.4	0.154	0.222
	organisation	0.656	0.459	0.54
	person	0.932	0.941	0.936
	overall	0.904	0.881	0.893

Named entity recognition: historical



INL labs α

Help

Home

Instituut voor Nederlandse Lexicologie

P.C. Hooft, Nederlandsche Historien · dbnl

title: P.C. Hooft, Nederlandsche Historien · dbnl

Named entity types: **Persons**, **Locations**, **Organisations**, **Misc**

LOUIS BOISOT (24)

Boisot (BOISOT) (9)

Louis Boisot (LOUIS BOISOT) (2)

Louis (LOUIS) (13)

VAN CHAMPAIGNEY (19)

van Champaigney (VAN CHAMPAIGNEY) (3)

Champaigney (CHAMPAIGNEY) (16)

LUIDEWYK (12)

Luidewyk (LUIDEWYK) (12)

DON SANÇIO (7)

Don Sançio (DON SANÇIO) (7)

CHIAPPIN VITELLI (6)

Chiappin Vitelli (CHIAPPIN VITELLI) (3)

Vitelli (VITELLI) (3)

GODT (6)

Godt (GODT) (6)

VAN GLIMES (5)

Glimes (GLIMES) (3)

van Glimes (VAN GLIMES) (2)

AVILA (5)

Avila (AVILA) (5)

JULIAAN ROMERO (5)

Juliaan Romero (JULIAAN ROMERO) (3)

Juliaan (JULIAAN) (2)

- Want de klem des bewinds, en 't oorlooghsvolk, was **Juliaan Romero** bevoelen.
- Daar onder was de vendrigh van **Juliaan**; en 's Kornels geluk, dat de brandt niet een ooghenblik laater opging: gemerkt hy zelf gereedt stond om oover te stappen; zoo de ridder **Williams** getuight, die zich toen in

[1574]

+MEt den aanvank des jaars vyftienhondert vierentseeventigh, stelde zich de nieuwe Landtvooghd in standt om den last der dingen t'onderscheppen, en handt aan werk te slaan. Twee hoofstukken, booven andere, waaren, ten afscheidt zyns voorzaats, onvoltrokken blyven hangen: 't beleg van **Leyde**, en 't ontzet van **Middelburgh**. Oordeelende +dat de zorgh des bewaarens, zonderling van zoo een stadt, voor moest gaan, dewyl het een en 't ander teffens niet te begaapen was, wendd' hy herwaarts zyn' gedachten, en quam in persoon t'**Antwerpen**, om, met zyn' jeeghenwoordigheit, +der toerustinghe scheut en hitte te geeven. Geen minder zorghvuldigheit dreef den Prins van **Oranje**, zich in persoon van **Delft** naa **Vlissinge** te vervoeghen; daar hy 't meeste vermooghen te waater van **Hollandt** en **Zeelandt** by een rukte. Don **Louis** bezeffende dat eene vloot, hoe geweldigh ook, in de engte der stroomen, met maar maatyke krachten, te stuiten waar, maakterder twee te gelyk reedt; d'eene, om, van **Antwerpen**, ter Honte af, d'ander', om, van **Berghen** op Zoom, ter **Scheld'** uit, te loopen; en in dier voeghe de maght zyner weederparthye te splitsen. Ende liet hy zich voorstaan, zoo zy beide de ruimte kreeghen, en gezaamentlyk den vyandt onder de ooghen quaamen, dat lichtelyk der eene oft der andre gelukken zoude, tot aan **Middelburgh**, oft Armuide, 't welk ook noch voor den Konink hield, deur te booren. D'Antwerpsche, ten getaale van dertigh zwaare scheepen, steld' hy onder den burghvooghd **Don Sançio**. Aan die van **Berghen**, bestaande uit tseeventigh smakzeilen, de voorraadtvoerders daaronder begreepen, gaf men, mits Beauvois dat pas in zyn doodbed lagh, den Heer **van Glimes** tot Ammiraal; meer om welstaans wille, en de Landzaaten te vlaayen, dan anderzins. Want de klem des bewinds, en 't oorlooghsvolk, was **Juliaan Romero** bevoelen. 'T zy dat men den Walschen Eedelling, hoewel vermaardt voor eenen braaven krysmen, die bequaamheit niet toevertrouwde, oft dat de Spaansche hooft niet ooverhoofd, dan van haaren landaart, lyden kon. Oover al de hondert scheepen werden drie welgevulde regementen verdeelt: naamelyk dat van **Beauvois**, van **Hierges**, en van **Romero**; 't welk vyftieh vendelen ouwde Spaansche soldaaten uitbraght. In 't laatste van Louwmaand; en vast op eenen tydt, ging **Avila** t'zeil, en Don **Louis naa Berghen**: van waar voort ook Glimes en **Romero**, den neeghenentwintighsten, afvoeren, met ongelukkig voorspoek. Want een van de busschieters, als hy den Landtvooghd ter eere, vuur gaf, had zich zoo

INT sequence taggers

Taggers

- PoS (modern/historical)
- NER

Technical details

- SVM^{light} (c++) for training
- Tagging in java
- Historical lexicon used for lemmatization, in combination with approximate match
- TEI input and output possible
- Uses word embeddings and (for NER) other semi-supervised features
- Currently published only as web services
 - training procedure invokes executable for SVM-light, rather involved

But....

This is all very nice but:

- Tools perform badly for non-standard-modern language
- Developing training material for other material is a huge effort

Required effort: training/eval data

To do: harmonize, share, make benchmark sets

Corpus Gysseling	1.6M
CRM	650k
Adelheid training sets	?
Gentse Brieven	70k
Brieven als Buit	400k
Hoofd letters	?
Statenvertaling	?
“Corpus Vosters”	~100k
IMPACT evaluation corpora	~20k
Nederlab evaluation data (tbd)	At least 10k per century

Tagset issues

CGN/DCOI dominant for modern Dutch, but some issues with historical Dutch:

- Principle *tag + lemma* \Rightarrow (almost) determines word form cannot be maintained historically
- (likewise for CG/CRM): tagging inflection instead of case/number/mood/tense etc does not benefit (diachronic) analysis
- Details:
 - Mood cannot be lumped with tense historically (PVTIJD = tegenwoordig, verleden, imperatief, conjunctief.)
 - “dial” for all non-standard-modern words is unsatisfactory

Diachronic and synchronic tagsets

- CGN/DCOI <http://www.hum.uu.nl/medewerkers/p.monachesi/papers/vaneynde.pdf>
- Corpus Gysseling/Corpus van Reenen-Mulder
- Adelheid tagset http://adelheid.ruhosting.nl/download/Adelheid1_TagsetManual.pdf
- Main PoS only in Brieven als Buit en Gentse Spelen
- GiGaNT tagset
http://www.ivdnt.org/images/stories/onderzoek_en_onderwijs/publicaties/TaalbankWorkingpaper3.pdf
- International standards (Gold, eagles, parole, ...)
 - Universal dependencies, adopted in CLARIN/CLARIAH
<http://universaldependencies.org/>

Smarter software?

Can we avoid a huge effort of training corpus building??

- More advanced supervised learning (deep....)
- Domain adaptation techniques
- Semi-supervised techniques
- Map to “modernized” Dutch
- Derive lexica from parallel text
- Transfer tagging by alignment

Tools for manual annotation

<https://github.com/INL/COBaLT>

CoBaLT is an application in which a corpus of texts can be loaded so as to be able to annotate its tokens (lemmatize and more). The annotation work in CoBaLT gives two products: an annotated corpus, and a lexicon consisting of word forms and their corresponding lemmata and such assigned to them.

<https://github.com/INL/AttestationTool>

The Attestation Tool is a multi purpose GUI used in the production of computational lexica and gold standard data for NE tagging. It addresses the situation of having a comprehensive dictionary consisting of headwords (lemmata) and corresponding quotations in which occurrences of the headwords were automatically matched.

Firefox CoBaLT - IMPACT Corpus-Based Lexico...

cobalt.inl.loc/lexiconTool.php?sDatabase=LexiconTool_NL_BUB&iCorpusId=5&sCorpusName=Brieven Als Buit&sUserName=tom&iUserId=5

word form frequency Go to wordform... Filter wordforms... Filter lemmata... 100 15 normal < Corpora << Start page

Prev 1 101 201 301 401 501 601 701 801 901 1001 1101 1201 1301 1401 1501 1601 Next

aansteekende	1	aanstekend, ADJ	aanstekend, ADJ
aanstellen	1	aanstellen, VRB	aanstellen, VRB
aanstelling	1	aanstelling, NOU	aanstelling, NOU
aanstipping	1	aanstipping, NOU	aanstipping, NOU
aanstonds	3	aanstonds, ADV	aanstonds, ADV
aanstonts	1	aanstonts, ADV	aanstonts, ADV
aanstotelijke	1	aanstotelijk, ADJ	aanstotelijk, ADJ
aant	4	aan, ADP & het, ART	rechtaan, ADV aan, ADP & het, ART aan, ADP & het, PRN
aantal	5	aantal, NOU	aantal, NOU
aante	2	aandoen, VRB & te, ADP aangaan, VRB & te, ADP	aandoen, VRB & te, ADP aangaan, VRB & te, ADP
aanteelen	1	aantelen, VRB	aantelen, VRB
aantegaan	1	aangaan, VRB & te, ADP	aangaan, VRB & te, ADP

This word was viewed last by Marijke at Mon 30 Jul 2012, 09:16:50

uploadedDocuments/06-01-2010 185-188-TR-defTanja.xml_tokenized.tab

kinderen dat met de nodige slavinne tot oppassing een aantal van somtyds tachtig personen maakt tans verhuyst aantal, NOU ✓

zo wel als t geheele huyshouden dat dan een aantal van 25 a 30 personen maakt, dog haare aantal, NOU ✓

uploadedDocuments/1330-1333-TR-def.xml_tokenized.tab

alzoohier een formidabele Rheede is met een groot aantal scheepen van alle naties zonder onderscheid, inzonderheidt aantal, NOU ✓

uploadedDocuments/5-1-2008 105,108-109-TR-def.xml_tokenized.tab

het hier by de meeste en Zelvs de gegoedste geleegeen aantal van Producten en wynig scheepsgeleegentheid. Zeer verwondert aantal, NOU ✓

uploadedDocuments/Vliet 54.xml_tokenized.tab

Campen om naar Guinea te komen, versien met een goet aantal van lantsoldaten, onder 10 orangie vandelen , alle aantal, NOU ✓

Named entity annotation tool=dictionary attestation tool

