Language Processing of 17th Century British English and Creation of a Diachronic Nooj Dictionary

Hélène Pignot
Senior lecturer in English, research field: English travellers in Greece in the 17th century
University of Paris I

Odile Piton
NLP Laboratoire Marin Mersenne
University of Paris I
Corpus: Accounts of British travellers in Greece and Anatolia in the 17th century

- **Authors**
  - George Sandys, Henry Blount, John Ray, Paul Rycaut, Thomas Smith, George Wheler,
  - 124000 tokens

- focus on Greek society

- Greece was part of the vast Ottoman Empire which spread across three continents
  - historical,
  - geographical,
  - sociological,
  - religious, (religious customs of the Greeks)
  - multilingual aspects: at least 36 different peoples

  → Hence a very rich lexicon.
From 17th Century English to Modern English

1. Record the evolution of POS

2. List changes in spelling, in meaning, in structure

3. Describe and transcribe all these changes with NOOJ Graphs + dictionary

4. Validate the entries in a NooJ dictionary
I. differences in parts of speech

1. General remarks for 17th English:
   Changes in spelling: double consonants, different prefixes and suffixes
   mute e appended or inserted in the word, u an w two forms of a single letter
   Inversions: infinite numbers there are
   Compound words / instead of single forms: for ever, every where

2. Articles: use of an for a before an aspirate h: an holy amulet

3. Adjectives: Some changes in word order: adjective after the noun Church-Christian

4. Conjunctions: though is spelt tho'

5. Prepositions: through is spelt thro'
   Thrusting an iron stake thro' the body out under the neck (torture of empalement!)
   prepositions are often omitted
   tho' dismiss the seraglio → though dismissed from the seraglio
6. Personal pronouns

- Spelling
  - *He, she, we, you, they:* *hee, shee, wee, ye, thay*
  - *Its* appears during the 17th century

- Forms
  
  Reflexive pronouns: *it self(e), my self(e) himselfe*

- contraction
  - *‘tis for it is*

- Use of the 2nd person singular (often used in prayers)
  
  nominative: *thou*  
  accusative: *thee*  
  genitive: *thine*  
  possessive adj: *thy*
7. The noun : spelling

• **hyphenated forms instead of concatenated ones**
  Sun-set, fore-finger, whirl-wind, Archi-pelago for sunset, forefinger, whirlwind, Archipelago
  ... Or on the contrary : **Northwind** for North wind

• **Different uses of 's :**
  1. 's is used as the interlingual plural of foreign nouns
     – *janizaries* and Spahi’s
     – *Bassa’s (Pashas), grotta’s, gondola's*
  2. The Saxon genitive (‘s) is used
     – *a Bassa’s son* by a Sultana
     – *The grand Signior’s women*
     – *in S. Peter’s Church*
  3. But most of the time we find the forms without the ‘:*
     *women’s, children’s, men’s, another's*
     - *from the women’s apartment*
     - *out of their children’s mouths*
7. Noun structures: in order to; xxx soever

- **Noun structure ‘In order to’**
  - modern English: ”In order to” + <V>
  - 17th English allows: ”In order to” + NP :
    - to beg their alms *in order to their relief*
    - *in order to his recovery*

  → In modern English: replace *in order to* by *for*

- **structure ‘xxx soever’**
  - at what <A> <N> soever

  → however <A> the <N> is

  \[ \text{at what great distance soever} \]

  → however *great the distance is*
7. Noun structures with soever

no <N0> of what <N1> soever

→ whatsoever the <N1> of a <N0>

no woman of what quality soever

→ whatever the quality of a woman

How <A> soever this <N> is

→ however <A> the <N> is

How strict soever this Church is

→ however strict this Church is
8. Verbs

- **Use of 2\textsuperscript{nd} person singular; 3\textsuperscript{rd} person PR**
  - *thou satisfyst*  
    *Hee, shee cometh, hath*

- **Preterite and past participle:**
  - Elisions inside the verb form ‘d for ed: *I observéd*
  - t instead of ed after phonems like k, sh, x:
    *if their churches were once demolisht*

- **Subjunctive:** very common in 17\textsuperscript{th} century English after
  (unless | lest | if | provided ) + (bare infinitive | simple PR as in ME)

provided it be done after a due manner \rightarrow provided it is done after a due manner
8. Verbs: present and preterite: two forms

• Besides the *usual forms*, one can use
  – do(e) + bare infinitive = present
  – did + bare infinitive = preterite

the do auxiliary is of course still used in modern English in the negative form

  do not + bare infinitive               did not + bare infinitive

• Present
  
  “the Greeks *doe now for the most part imitate* the Turks (I mean here in Turkie) in sitting at their meal” (Sandys)
II Lexicon

• Toponyms: *Candy for Crete*

• Ethnonyms: *Freinks or Franks for western Christians*

• Historical Terms:
  - names of currencies: *zecchines* = Venitian coins (sequins),
    *hungar* = Hungarian coin
  - Units of weight: *Oque* = 1kg 250  Unit of distance: *League* ca 4 km

• Names of Lay or Religious Titles borrowed from other languages:
  - Turkish: *Bassa* = the Pasha
    *Keslar-Agafi* = a black eunuch who watches over the ladies in the Seraglio
  - Greek: *Egoumen* = the Abbot of a monastery
  - *vroukolakas* = an evil spirit
  - Italian: *a capriccio of the Grand Signor* = a caprice of the Sultan
II Lexicon

• Words from other languages used in the text to add local colour

randevouzes: the English plural of the French word rendez-vous!
salam’d : ‘salamed’ = saluted English form of an Arabic word
mamoukode = a ghost
Kaloir: from the Greek Kalogieros = a good elder: a Greek Orthodox monk
tó déma = tying up a man from accompanying with any woman
= a spell to make a man impotent (in French= nouer l’aiguillette)
Odaliques = odalisques, ladies of the Seraglio
Kara Congia = a demon appearing in the shape of a black old man
II Lexicon: word evolution + expressions

• There are interesting semantic changes:
  * ghostly = spiritual    miscarriage = misbehaviour

• archaic words or expressions:
  * Turcism = Islam
  * Betimes = early
  * Prejudicacy = prejudice
  * to symbolize with someone = to resemble
  * a penitentiary = can be a spiritual father or a penitent!

• words sometimes derived from Latin or Greek
  * flagitious (from the Latin flagitiosus) = guilty of terrible crimes
  * ethnick (from the Greek ethnos) = pagan

→ entries in the NOOJ dictionary
III- Graphs: diachronic computable variations

Method
1. to list morphological transformation rules
2. to build graphs to recognise archaic forms dynamically
3. to validate the results
4. to build entries for a new dictionary

(+ to use graphs for computed words)

We are presenting some of our graphs and discussing the results here
III 1-list morphological transformation rules

• A- Prefixes / Suffixes
  - transformation **ncy** → **nce**
    - **observancy** → **observance**

    ous for ate : **degenerous** for **degenerate**
    ship for hood : **priestship** for **priesthood**

• transformation **en** → **in**, **em** → **im**:
  - **increasing** for **increasing**, **emboss** for **imboss**
III- Graphs: diachronic computable variations

B- Different spellings:

mute e : final or not  
   doe, losse, onely for do, loss, only
omitted e :  
   entred, wherof for entered, whereof
double letters inst. of single letter :
   equall, chappel, shooe for equal, chapel, shoe
y for i, w for u, i for j :  
   oyl, coyn, lew for oil, coin, Jew
ick for ic :  
   Arabick, garlick for Arabic, garlic,
eer for ear :  
   yeer for year
edge for ege :  
   alledge for allege
ence for ense :  
   expence for expense
B- Spellings

This graph spots the transformation \textit{double letter} \rightarrow \textit{single letter}

\[ <\text{MOT1}<\text{LET}<\text{LET}<\text{MOT2}>_{\text{DIC}}> \]

\[ e.g.: \text{sollicitous} \rightarrow \text{sollicitous} \]

1. The debug shows that \textit{sollicitous} is decomposed in the sequence \texttt{MOT1\_LET\_LET\_MOT2} With MOT1 = so LET = l MOT2 = icitous

2. The sequence \texttt{MOT1\_LET\_MOT2 = sollicitous} is compared to the dictionaries,

3. \textit{sollicitous} is recognised as an Adjective,

4. the debug displays in green color: <LU=sollicitous,A>
B- Spellings

The same graph works for

\[ \text{bigottry} \rightarrow \text{bigotry} \]

that is recognised as a noun

Enter expression: \text{bigottry}

Paths

\[ ("\text{Main" } \\langle \text{MOT1 bigo } \rangle \langle \text{LET t } \rangle \langle \text{MOT2 ry } \rangle ) \langle \text{LU=bigotry, bigotry, N+FLX=ACTIVITY+Nb=s} > \]

Outputs

shooes \rightarrow \text{shoes}

that is recognised as a noun, and as a verb

Enter expression: \text{shooes}

Paths

\[ ("\text{Main" } \langle \text{MOT1 sh } \rangle \langle \text{LET c } \rangle \langle \text{o}(\text{MOT2 es } \rangle ) \langle \text{LU=shoes, shoe, N+FLX=APPLE+Nb=p+Distribution=Conc} > \]
\[ ("\text{Main" } \langle \text{MOT1 sh } \rangle \langle \text{LET c } \rangle \langle \text{o}(\text{MOT2 es } \rangle ) \langle \text{LU=shoes, shoe, V+FLX=SHOE+Tens=ER+Pers=3+Nb=s} > \]
B- Spellings

transformation \( \text{ence} \rightarrow \text{ense} \) at the end or in the middle of a word

sence \( \rightarrow \) sense

sense is recognised as a noun and as a verb

expencefull \( \rightarrow \) expenseful

But expenseful is not recognised for two reasons (see later)
C- Combination of several variations recognises *wasteful* from *wastfull*: add mute e, delete one l

**This graph**
- builds *wasteful* from *wastfull*,
- compares it to the *dictionary*
- recognises it as an *adjective*
C- combination

recognises *won* from *wonne* : delete one n

```
<init> ( <L> ) → ( <L> ) LET m1 e <s1$LET=:DIC
```

<table>
<thead>
<tr>
<th>Paths</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&quot;Main&quot; $(m1 wo $) $(LET n $) ne )</td>
<td>&lt;LU=won, win, V+FLX=WIN+Tense=PT+Pers=1+Nb=s&gt;</td>
</tr>
<tr>
<td>(&quot;Main&quot; $(m1 wo $) $(LET n $) ne )</td>
<td>&lt;LU=won, win, V+FLX=WIN+Tense=PT+Pers=2+Nb=s&gt;</td>
</tr>
<tr>
<td>(&quot;Main&quot; $(m1 wo $) $(LET n $) ne )</td>
<td>&lt;LU=won, win, V+FLX=WIN+Tense=PT+Pers=3+Nb=s&gt;</td>
</tr>
<tr>
<td>(&quot;Main&quot; $(m1 wo $) $(LET n $) ne )</td>
<td>&lt;LU=won, win, V+FLX=WIN+Tense=PT+Pers=1+Nb=p&gt;</td>
</tr>
<tr>
<td>(&quot;Main&quot; $(m1 wo $) $(LET n $) ne )</td>
<td>&lt;LU=won, win, V+FLX=WIN+Tense=PT+Pers=2+Nb=p&gt;</td>
</tr>
<tr>
<td>(&quot;Main&quot; $(m1 wo $) $(LET n $) ne )</td>
<td>&lt;LU=won, win, V+FLX=WIN+Tense=PT+Pers=3+Nb=p&gt;</td>
</tr>
<tr>
<td>(&quot;Main&quot; $(m1 wo $) $(LET n $) ne )</td>
<td>&lt;LU=won, win, V+FLX=WIN+Tense=PP&gt;</td>
</tr>
<tr>
<td>(&quot;Main&quot; $(m1 wo $) $(LET n $) ne )</td>
<td>&lt;LU=won, won, N+FLX=SERIES+Nb=p+Distribution=Unit&gt;</td>
</tr>
</tbody>
</table>
C- combination

recognises mere from meer: delete one e, add one e

This graph
• builds mere from meer
• compares it to the dictionary
• recognises it as an adjective and as a noun (old word = border)
• recognises it falsely as « mer »: mer is not an English word, it is an abbreviation for meridian

→ each proposition has to be validated
D conjugations : PR 2th and 3rd person

• Recognises the second person dynamically
  *thou hast (to have), thou art (to be) → dictionary*
  *thou comest (P), thou camest (PR) to came → graph*

• Recognises the third person in th dynamically
  – desireth desires : eth → es  *when he most desireth*
  – differeth differs : eth → s
  – doth does : th → es  *hee doth not much care*
  – saith says : ith → is  *four different graphs*
D conjugations preterite and PP

- Forms with t: s, f: double the consonant + ed; h, k, x: add ed, p either double or not + ed
  - confest: confessed
  - linkt: linked
  - fixt: fixed
  - cropt: cropped
  - quaft: quaffed
  - establisht: established
  - ravisht: ravished
  - helpt: helped

- Disambiguate the forms in ‘d.
  - I observ’d abundance
  - we observ’d firrs

- ‘d can appear in a (compound) adjective: the long-hair’d Greeks

24
... when Antipater, Perdicas, Seleucus, Lysimachus, Antigonus, Ptolemey, and the rest of the successors of Alexander had shared his empire among them, they endeavoured as much as they could to plant his new-got kingdoms with their countrymen: whose posterity in part remaineth to this day, (though vassaled to the often changes of foreign governors: supplied by the extension of the latter Greek Empire; who yet retain wheresoever they live, their name, their religion, and particular language. A nation once so excellent, that their precepts and examples doe still remain as approved canons to direct the mind that endevoureth virtue. Admirable in arts and glorious in arms; famous for government, affectors of freedom, every way noble: and to whom the rest of the world were reputed barbarians.
Discussion

• Generally speaking dynamic word formation in 17th century English is very close to modern English, a NooJ graph would be very useful.

• A word like expenseful is not recognised because it’s not in the NooJ dictionary.

• Problem of multiple modifications: endevoureka, endeavours.

• Lemmas + transcription difficulties with short words: with some transformations (add e, change letters...), they look like another word → the graphs can suggest wrong interpretations

  The Greek word sas was transformed into seas

  → each proposition has to be validated
Conclusions

• Use of “NooJ translation” but specificity of transcription from 17th English to modern English
• Spelling is often phonetical
• But many constructions which were used in the 17th are standard English today → 17th century syntax is very close to modern English
• Nooj graphs help their users to transcribe 17\textsuperscript{th} century English into modern English
• NOOj is a superb tool for linguists working on 17\textsuperscript{th} century English texts. Nooj has great pedagogical potential: it may be used by students and teachers in class (to compare 17\textsuperscript{th} century English with modern English)
Excerpts about Hungary

• ...the hatred of the Greeke Church to the Romish was the losse of Belgrado in Hungary,...
• ...they avoid quarrelling with two Christian princes at once, being usually at league either with Poland and Muscovy, when they war upon Hungary, and so on the contrary; dreading nothing more than a union of the Christian princes bordering upon them...
• This they are very sensible of, and therefore as they take all occasion to promote quarrels and dissensions in Hungary and Transylvania, so they greatly rejoice, when the princes of christendom are at war one with another.
• ...couriers are dispatch’d from Constantinople to Ragotski, prince of Transylvania, ...
• ...in Transilvania.
• Hence it comes to pass, that zecchines and hungars for gold, and Spanish dollars and zalotts for silver stamp’d in christendom ...