HAL
open science

LANGUAGE PROCESSING OF 17 TH CENTURY ENGLISH WITH NOOJ

Hélène Pignot, Odile Piton

To cite this version:

Hélène Pignot, Odile Piton. LANGUAGE PROCESSING OF 17 TH CENTURY ENGLISH WITH NOOJ. 11th NooJ conference and workshop, The Department of Language Technology of Research Institute for Linguistics (Hungarian Academy of Sciences), Laboratoire de Semio-Linguistique, Didactique et Informatique of Université Franche-Comte, and Maison des Sciences de l'Homme Claude Ledoux, Jun 2008, Budapest, Hungary. pp.191-203. hal-01091219

HAL Id: hal-01091219

https://hal.science/hal-01091219

Submitted on 4 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LANGUAGE PROCESSING OF 17TH CENTURY ENGLISH WITH NOOJ

HÉLÈNE PIGNOT AND ODILE PITON,
UNIVERSITY OF PANTHÉON-SORBONNE

Abstract: In this article we have explored ways of using the NooJ software to describe the differences in spelling, meaning, and grammar between 17th century English and contemporary English. We based our study on a corpus of six 17th century writers who travelled in Greece and Anatolia.

Keywords: 17th Century English, NLP, dynamic processing, electronic dictionary, morphological analysis, morphological graph, syntactical grammar**.**

Introduction

Reading the accounts of 17th century English travellers in Greece and Anatolia may seem a daunting task for non-Anglophones. First one has to be familiar with the historical and geographical context of these travels. Greece at the time was part of the vast Ottoman Empire which spread across three continents. Second, mentalities were extremely different, thus for 21st century readers these tales (which were best-sellers in their day) imply not only a change of scene but a tremendous change of time and perspective. Last but not least, not only does the content of these tales sound strange and intriguing, its form may also strike the modern reader as singular and extremely varied.

Indeed 17th-century spelling includes many variants and the reader may come across the same word spelt in three different ways in the same book. Some words have fallen into disuse; others have taken on different meanings today. This variety and this singularity do not lack charm, but while reading these texts I wondered if one might describe all these variants and discover any regular patterns that could account for them? Computing them systematically and using NooJ graphs to describe them could help readers—especially non-Anglophones—make sense of their form and their meaning.

Our corpus comprises excerpts from the accounts of six British travellers in Greece and Anatolia in the 17th century (in chronological

order): George Sandys, Henry Blount, John Ray, Paul Rycaut, Thomas Smith, and George Wheler[1]. This corpus generated 124,000 tokens. Our authors' focus on Greek society includes historical, geographical, sociological, religious and multilingual remarks: hence a very rich and varied lexicon with Greek, Latin, Turkish, Italian, Spanish, French and English words.

I. Word classes.

1. 1. General introductory remarks.

There are many differences in spelling such as double consonants, different prefixes and suffixes and mute "e"s appended or inserted into words. U and w, i and j are two forms of a single letter. Some adverbs are written as two lexemes instead of one compound word, as in "for ever" or "every where". Punctuation is different too. The colon is used to mark either the end of a sentence or a pause in a sentence (especially in Sandys). The ampersand (&) often replaces the conjunction "and".

1.2 Lexical categories

♦ Nouns and pronouns: forms and noun structures.

Many words are hyphenated instead of concatenated (for example sun-set, fore-finger, or Archi-pelago) or vice-versa, hence the North wind is spelt Northwind.

There are various uses of the inflection /'s/: besides the genitive case —as in *the grand Signior's women*—, 's is used as the interlingual plural of foreign nouns as in *Spahi's, Bassa's (Pashas),* or *gondola's.* The genitive case may be spelt without the apostrophe (a form closer to OE): "*from the womens apartment"; "out of their childrens mouths".*

The following noun structures are noteworthy: **no <N1> of what <N2> soever** and **in order to + N=** for**,** as in *in order to their relief* (these structures are analysed in part III).

---

[1] George Sandys, p 60-66; H. Blount, p 108-113; John Ray, p 7-11; Paul Rycaut. p 1-21, 120-38, 161-70, 177-87, 199-200, 293-317, 321-36, 375; Thomas Smith, p 14-19, 27-30, 51-3, 62-78, 79-80, 90-8, 158-62, 178-202, 211-3, 217-21, 230-3; George Wheler, p 194-99, 245-49.

♦ **Personal pronouns and adjectives.**

The alternative spellings hee, shee, wee, ye, thay may be found instead of he, she, we, you, and they. Reflexive pronouns are not single lexical items but two written words as in *it self(e), my self(e), him self(e).* The 2nd person singular pronouns are often used (especially in prayers to God and the saints): e.g. "*We offer to thee thine of thine own, in all things and through all things*" (Smith).

As for adjectives, some variations in word order may be noted; the adjective can be placed after the noun (e.g. "*their difference theologicall I enquired not*"). Adjectives may be used as adverbs: "*they are wonderfull curious in the composition of it*".

♦ **Verbs: morphology and verb paradigms.**

In the present, two forms are noteworthy: the 2nd person singular, with the inflectional morpheme – est, as in *thou satisfyest* and the 3rd person singular, with the inflectional morpheme – eth, as in *shee cometh, he hath, he endevoureth.*

In the preterit, some verbs that are irregular in modern English had an inflectional ending in – ed (exs: catched is the preterit of catch, shined the preterit of shine). Others had different forms, for instance to speak (I spake) or to begin (I begun, instead of began in CE).

• Declarative, interrogative and negative forms.

a. In the declarative form of the present and the preterit, besides the verb (with the inflection "s" in the third person singular), the auxiliary do(e) with the bare infinitive may be used: "*They [the Greeks] are of diverse trades in cities, and in the country do till the earth (for the European Turks do little meddle with husbandry)*" (Sandys). *"Mens minds did labour with fearefull expectations"* (Sandys). Do is also used in the imperative: "*do thou bless this oblation*", "*do thou pardon, as being our good and gracious God*" (Smith).

b. In the interrogative form, the auxiliary verb "do" or the verb with an inverted subject are also used. These accounts do not contain any dialogues. Reading other 17th century texts which include dialogue, such as plays, we can see that questions did not always require the auxiliary verb "do", as in old English. Let's give two examples taken from Aphra Behn's plays:

> "What mean you by this language?" (*The Dutch Lover*). "How came she in?" (*Abdelazer or the Moor's Revenge*).

c. In the negative form besides the auxiliary “do” + not + verb, one also finds the structure verb + not: as in *“the women marry not till the age of 24”* (Sandys), or *“their difference theologicall I enquired not”* (Blount).

♦ **The subjunctive.** The subjunctive (i.e. the bare infinitive for all persons) is very common. It appears in conditional and concessive clauses (whereas in contemporary English it occurs chiefly in formal contexts). It is used with unless | lest | if | provided | though | whether | whatever, as in *“provided it be done after a due manner; “ “if they be kaloirs”* (Smith), “*whatever it be*” (Smith), but also in temporal clauses, *“until they be eight years of age”* (Rycaut), and comparative clauses, “*they had rather remaine as they be*” (Sandys).

♦ **Prepositions.** Through may be spelt thro’ (e.g. “thrusting an iron stake thro’ the body out under the neck”). Prepositions are often omitted: “*but our subject here is more tragical, the subversion of the sanctuaries of religion, the royal priesthood expelled their churches, and those converted into mosques*” (from is omitted).

1. 3. Functional categories.

a. Determiners. Usage is very close to CE except for that of the indefinite article “an” before words starting with h, which shows that the h was dropped (e.g. *an holy amulet).*

b. Complementizers: conjunctions and relative pronouns. *Though* may be spelt *tho’*. The pronoun “which” may be used when the antecedent is personal: *“The Georgians, which in some manner depend on the Greek Church, baptize not their children until they be eight years of age”* (Rycaut)

c. Wh-compounds: the forms whosoever, wheresoever and whensoever are commonly used, as well as the structure how <A> soever this <N> is : “*How strict soever this Church is esteemed in admitting many degrees of marriage*” (Rycaut).

All of these differences may be described in a NooJ grammar of 17th century British English. As for semantic changes they may be indexed in a NooJ dictionary.

II. Lexicon.

Ethnonyms and toponyms (which we have indexed in a NooJ dictionary) were often different in 17th century English. Their spelling also varied. Let us give a few examples. The word “Freinks” or “Franks” designates western Christians; “Hollanders”, Dutchmen; “Grecians”, Greeks; “Zynganaes” or “Zinganies” are Oriental people or gipsies.

“Candy” (Candie, or Candia) is the other name of Crete; “Roma(g)nia” designated Greece and the Balkans. The Mani (the South of the Peloponnese) is called the “Maina” and Croatia, “Sclavonia”. The “Arches” are the islands of the Archipelago, i.e. of the Ægean Sea, between Greece and Asia Minor.

Table 2.1. Historical Terms

Category	Example	Definition
Currencies	*A zecchine*	a Venitian coin (sequin)
	A hunga	a Hungarian coin
	A sultany	a Turkish gold coin
Units of weight	*An oque*	1kg 250
Unit of distance	*A league*	ca 4 km
Turkish title	*Bassa*	the Pasha
	Keslar-Agafi	a black eunuch who watches over the ladies in the Seraglio
Greek title	*Egoumen*	the Abbot of a monastery
Italian title	*the Grand Signior*	the Sultan

Foreign words are also used in travellers’ accounts to add local colour and to describe Greek society which at that time revolved around religion. Travellers like Rycaut and Smith give us a unique insight into the rites and the events (festivals, fasts, and sacraments) that punctuate the life of the Orthodox Church, taking care in quoting the Greek text of the various rituals. They suggest translations and equivalents in English wherever possible. Some archaic words or meanings are “Englished” forms of Latin or Greek words.

Table 2.2. examples of loan-words

language	17th century Eng.	Contemporary English
French	*randevouzes:*	English plural of the French word rendezvous!
Arabic	*salam'd;'salamed'*	saluted
Greek	*tó déma*	*"tying up a man from accompanying with any woman"* a spell to make a man impotent
Greek	*mamoukode*	A ghost
Greek	*vroukolakas*	an evil spirit
Turkish	*alempena*	Constantinople or the refuge of the world (Smith)
Turkish	*kabin*	cohabitation as opposed to marriage
Turkish	*kara congia*	A demon appearing in the shape of a black old man
Italian	*a capriccio of the Grand Signior*	A caprice of the Sultan
Greek	*kosmokratores*	Emperors of the world (Smith), the Byzantine Emperors called themselves that way.
Greek	*sponsalia*	The betrothal ring (Rycaut)
Greek	*metousiosis*	transubstantiation[2]
Greek	*somatikos*	Corporally (Wheler)
Greek	*kaloir (kalogieros)*	A good elder: a Greek Orthodox monk
Greek	*paranomoi*	flagitious persons, and transgressors of the laws and canons of the Church (Smith)
Greek	*ethnick*	pagan, heathen (ethnos, OED 1470)
Latin	*flagitious*	guilty of terrible crimes (flagitiosus, OED, 1550)
Latin	*supposititious*	spurious (supposititius, OED, 1611)

There are interesting semantic differences that we indexed systematically. Here are a few examples of these archaic meanings, rare words or expressions:

Table 2.3. Examples of semantic evolution

17th century English	Contemporary English
a miscarriage	misbehaviour
Turcism	Islam
a mescheeto (or meskeeto)	is not a virulent insect but a mosque!
prejudicacy	prejudice
a penitentiary	a spiritual father or a penitent according to context. (an interesting case of enantiosemy**)**

[2] Transubstantiation: an application of the Aristotelian categories of substance and accidents to the sacrament of Holy Communion: after consecration, the substance of the wine and bread has changed to become the Body and Blood of Christ and there only remains the "accidents" of bread and wine. Rycaut remarks that this word is a recent invention of the Greeks, "which they never read in their ancient Fathers". So does Smith.

an ague	a high fever (OED, 1611)
an acquist	a purchase
ghostly	spiritual, e.g. “ghostly father”.
to symbolize with someone	to resemble, to partake of the nature of
to unintermix	not to mingle with so
to affect	to love

III Graphs: diachronic computable variations.

We have listed the rules of morphological transformations and used transducers to recognise unknown words dynamically. Some of our graphs are presented below and the results will be discussed further down. A detailed presentation may be consulted on our website:

http://marin-mersenne.univ-paris1.fr/site_equipe_mm/O_Piton

III.1-List the rules of morphological transformations

The transformations can be prefix, infix or suffix variations.

Table 3.1. Examples of morphological transformations

transformation	Samples	Transformation	samples
em for im	*emboss*	in for en	*intangling; intrench*
en for in	*encreasing*	in for un	*ingrateful* (arch)
un for in	*uncapable*	ie or y, for ey	*Turkie, Turky*
au for a	*launce*	eer for ear	*yeer, neer*
y for I	*oyl, coyn*	ea for e	*compleate, seaven*
e for ea	*endevour*	ei for ai	*straight*
ai for ei	*soveraigne*	n for gn	*forraine* for foreign
ee for ie	*peece*	edge for ege	*alledge, colledge*
ence for ense	*expence*	oa for o	*shoar, cloaths, choake*
ou for oo	*bloud*	our for or	*emperour, terrour*
w for u	*perswasion*	ie or ey for y,	*christianitie, countrey*
ph for f	*phantastique*	ment for tion	*deprivement* for deprivation
i for j	*iustice, Iew*	ick, ik for ic	*Arabick, garlick, traffik*
ship for hood	*priestship*	ncy for nce	*occurrency*
ous for ate	*degenerous*	que for c	*mathematiques, publique*

NooJ transducers may determine whether an unknown word can come about as the result of a morphological transformation. NooJ may check whether the prefix *in* may be replaced by ‘un’ or by ‘en’ (see fig. 3.1.). The debug shows that *encreasing* is correctly identified as a variant of *increasing*, and may be recognized as a gerund, a noun, or an adjective. *Uncapable* is recognized as the adjective *incapable*.

Fig 3.1. Morphological graph to change a prefix: un or en into in, and debug.

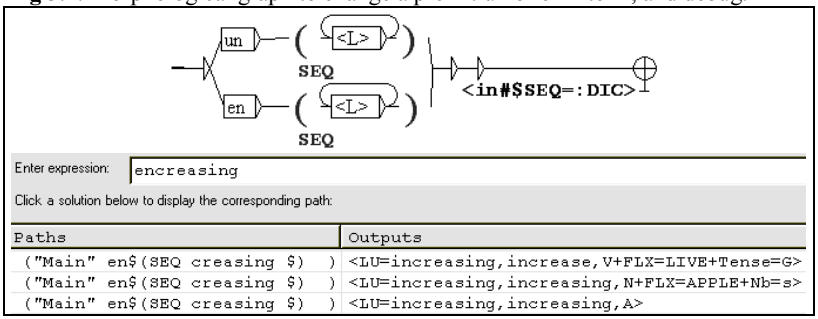


Figure 3.2. shows that an unknown word with the suffix ‘ncy’ may be recognised as a word with the suffix ‘nce’. *Occurency* is recognised as the noun *occurrence*.

Fig 3.2. Morphological graph to change the suffix ncy into nce.

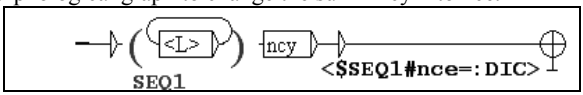
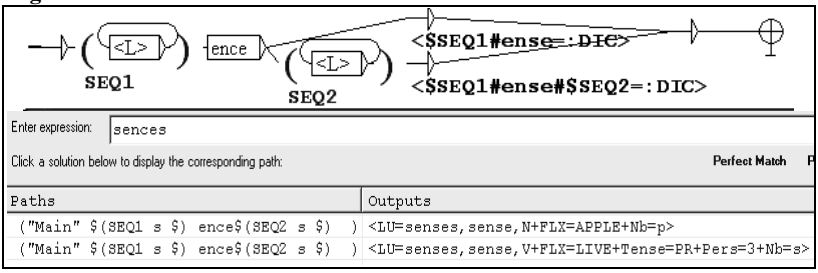


Figure 3.3 shows the transducer may recognise ‘ense’ for ‘ence’. Sences is recognised as a noun and as a verb. The same graph analyses expencefull, but expenseful is not clearly identified as it is not in the NooJ dictionary.

Fig. 3.3. Transformation of ence into ense at the end or inside a word.



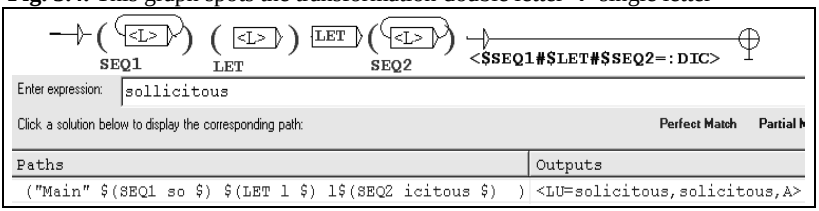
III.2- Insertion and deletion of letters

A mute /e/ may be added to the end of a word or inserted into a word as in “doe”, “onely”, or omitted inside a word as in “entred”. Consonants as well as vowels can be doubled as in *equall*, *lustfull*, *chappel*, *mysticall*, *shooe*, *bee* (be!), and *hee*.

Figure 3.4 shows a NooJ graph to recognise double letters instead of single ones. The word *sollicitous* is a good example: it may be broken down into the sequence SEQ1/LET/LET/SEQ2 with SEQ1 = so, LET = l, SEQ2 = icitous. The sequence SEQ1/LET/SEQ2 = solicitous is compared to the dictionaries, and the word is recognised as an adjective.

The same graph works for *bigottry* which is recognised as a noun (bigotry), and for *shooes* which is recognised as a noun and as a verb.

Fig. 3.4. This graph spots the transformation double letter → single letter



These are only limited examples. NooJ may combine greater graphs. See Fig. 3.5. This graph is applied to *wastfull*: it adds a mute e, deletes one "l" and recognises the word as an adjective.

Fig. 3.5. A combining graph adding a mute e and transforming a double letter into a single one

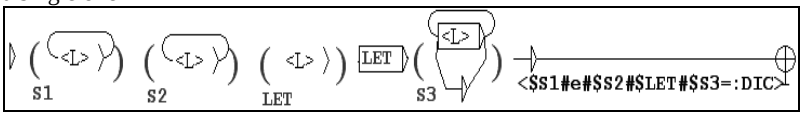
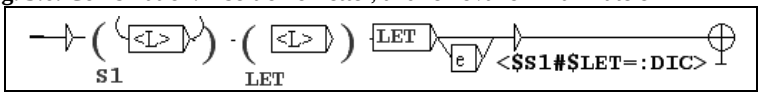


Fig 3.6. The graph is applied to *wonne*. It deletes one n and the final mute e, and then it recognises *won* as a verb and a noun. This graph recognises *mere*—from *meer*—as an adjective and as a noun (this old word means a border). It recognises also *mere* falsely as "mer": mer is not an English word, it is an abbreviation for meridian, so each proposition has to be validated.

Fig. 3.6. Combination: insertion of letter, and removal of final mute e

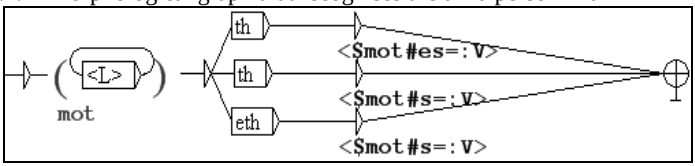


III.3. Transducers to recognise conjugations.

We have used transducers to recognise 17th century conjugations as "thou art" and "thou hast" are not in the NooJ dictionary. We have also

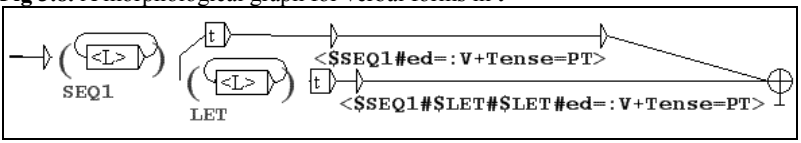
used NooJ graphs for the 17th century forms: e.g. thou *comest* (P), thou *camest* (PR) of the verb *to come*, as well as for the third person singular inflection in eth—as in *desireth* instead of *desires*, *doth* for *does* or *saith* for *says*.

Fig 3.7. A morphological graph that recognises the third person in th



Verbs ending with *ss*, *ff*: may be spelt either *ed* in the preterit or *t*. The double consonant becomes a single one as in *confest/ confessed; quaft/ quaffed*. As for verbs ending with p, k, x, sh: the preterit is spelt ed or t as in *crop/ cropped; linkt/ linked; establisht/ established; fixt/ fixed; ravisht/ ravished*. The graph in fig 3.8. recognises the variants of the preterit.

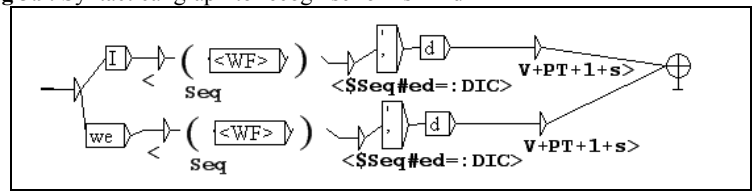
Fig 3.8. A morphological graph for verbal forms in t



III.4. Syntactical graphs

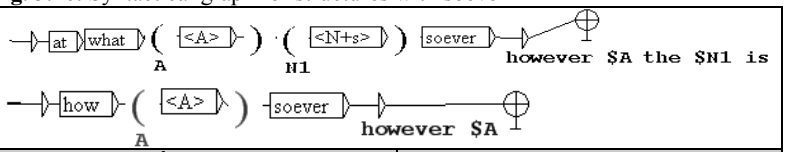
An apostrophe inside a word (as in the preterit) requires the use of a syntactical graph. This remark also applies to compound adjectives (e.g. *"the long-hair'd Greeks"*.

Fig 3.9. Syntactical graph to recognise forms in 'd



Structures with soever can be changed into the CE English form with however. See figure 3.10.

Fig. 3.10. Syntactical graph for structures with soever



17th English	Contemporary English:
How strict soever this Church is at what great distance soever	however strict this Church is however great the distance is

"In order to" is an interesting structure that may be followed by a noun in 17th century English. A graph easily replaces '*in order to*' by '*for'* e.g.: *in order to his recovery* becomes *for his recovery.*

The NooJ dictionary we have created records all validated results, including changes in meaning and the foreign words mentioned in the corpus. See Table 3.2.

Table 3.2. Some entries of our NooJ dictionary.

deprecation,N+FLX=Nsp+EN="intercessory prayer"+XVII
deprivement,N+FLX=Nsp+EN=deprivation+XVII
dissentaneous,ADV+EN="contrary to"+XVII
ethnick,A+EN=pagan+XVII
flagitious,A+EN="guilty of terrible crimes"+XVII
fœtid,A+EN=fetid+XVII
ghostly,A+EN=spiritual+XVII

Discussion and Conclusion

NooJs graphs compute words and compare them to its dictionary. The Nooj dictionary is not exhaustive and some words (such as expencefull) were not recognized.

A problem also arose with multiple modifications, for example "endevoureth" (endeavours in CE). While it is easy to compute one single morphological transformation, it is indeed much more difficult to deal with words that require several transformations. This includes additions and/or deletions of mute "e"s. The graphs must be applied sequentially. This gives rise to many transformations inside a single token which can produce wrong results. Thus a problem arose with the recognition and interpretation of double consonants: a word like farre (= far) was first recognised as a variant for fare or for fear! "Losse" was correctly recognised as the 17th century spelling of loss (with a redundant final e) but also indexed as a variant for lose! So each entry had to be checked, validated and included in a NooJ dictionary.

17th century spelling varies but these variations follow certain patterns as the graphs in part III show. Many structures used in the 17th century are standard English today, but some have disappeared. A NooJ grammar may record these structures and give their equivalent in Contemporary English. NooJ graphs help their users to transcribe and visualize these changes. A Nooj grammar and NooJ graphs are superb tools for linguists working on 17th century English. They have great pedagogical potential as they may be used by EFL students and teachers in class to compare 17th century English with modern English, hence facilitating access to 17th century literature for non-native speakers.

Bibliography

Blount, Henry, *A Voyage into the Levant* (London, 1636).

Pignot, H. *La Turquie chrétienne: récits des voyageurs français et anglais dans l'Empire ottoman au XVIIe siècle* (Vevey: Editions Xénia, 2007).

Pignot, H. *Comparaison entre l'anglais du XVIIe siècle et l'anglais contemporain à travers les récits de trois voyageurs britanniques.* Colloque " NooJ : outils pour la traduction automatique, quelles fonctionnalités développer ? ", Paris 1 Panthéon-Sorbonne, 29 septembre 2006, Sep 2006, Paris, France.

Pignot, H. *Christians under the Ottoman Turks : French and English Travellers in Greece and Anatolia (1615-1695)*. Piscataway: Gorgias Press, to be published in 2009.

Piton O., Lagji Kl. *Morphological study of Albanian words, and processing with NooJ.* Xavier Blanco, Max Silberztein (eds.), "Proceedings of the 2007 International NooJ Conference, Cambridge Scholar Publishing (2008), pp.189-205, Proceedings of the 2007 International NooJ Conference. <www.cambridgescholars.com>. <hal-00452458>

Ray, John, *A Collection of Curious Travels and Voyages: containing observations made by several learned and famous men in their journeys through the Levant* (London, 1693).

Rycaut, Paul. *The Present State of the Greek and Armenian Churches* (London, 1679).

Sandys, George, *A Relation of a Journey begun An. Dom. 1610* (1615; London, 1652).

Silberztein, M. *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX*. Masson Ed. Paris Milan Barcelone Bonn (1993)

Silberztein, M. "*NooJ's dictionaries"*, in Zygmunt Vetulani (ed.), "Proceedings of the 2nd Language & Technology Conference, April 21-23, 2005, Poznán, Poland"

Smith, Thomas, *An Account of the Greek Church*, London, 1680.

Wheler George, *A Journey into Greece* (London, 1682).