



HAL
open science

Nuptiae Philologiae... et X-Query

Christian Nicolas

► **To cite this version:**

Christian Nicolas. Nuptiae Philologiae... et X-Query. *Koinōnia* = Koinōnia, 2014, 38/2, pp.21-39. <hal-04620931>

HAL Id: hal-04620931

<https://univ-lyon3.hal.science/hal-04620931v1>

Submitted on 22 Jun 2024

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.



HAL Authorization

KOINΩNIA

38 / 2

2014

Christian Nicolas,

« Nuptiae Philologiae... et X-Query »

p. 21-40

È con molto piacere che faccio ritorno a Napoli per la chiusura del progetto didattico *Gli studi classici e la società dell'informazione* proposto da Marisa Squillante, per il quale l'Università Federico II ha ottenuto un importante finanziamento nel quadro del programma nazionale *Messaggeri della conoscenza*. Non stilerò un bilancio delle nostre attività ma piuttosto proporrò un elogio delle *Digital Humanities (DH)* con un'esortazione al loro utilizzo e mostrerò, servendomi in particolare dell'esempio del programma HyperDonat di cui sono il co-ideatore insieme a Bruno Bureau, qual è l'apporto delle nuove tecnologie ai nostri studi sull'antichità.

Comincio da una 'sorpresa retrospettiva': non avrei mai immaginato, otto anni fa, che sarei stato riconosciuto da parte del MIUR come esperto internazionale nel campo della filologia digitale. In effetti otto anni fa non avevo alcuna conoscenza teorica né, a maggior ragione, pratica dei metodi delle *Digital Humanities*; come molti dei membri della nostra comunità scientifica non credevo che alla carta stampata (e continuo a crederci) e consideravo l'informatica una comoda risorsa per organizzare l'impaginazione di un prodotto finale cartaceo. Conoscevo, apprezzavo e utilizzavo chiaramente i *database* che fanno guadagnare molto tempo allo studioso. Sapevo già che era inutile ricopiare un testo latino o greco dal momento che ne esisteva già sicuramente una versione su CD-Rom o sul web. E ignoravo che tali risorse, attualmente familiari a tutta la comunità, rientravano nelle *DH*.

1. Genesi del progetto

È stato durante un *brainstorming* del nostro centro di ricerca lionese che ho proposto un progetto sul commentario di Donato a Terenzio. Avevo letto questo testo nell'edizione Wessner (Teubner, 1902) e l'avevo trovato strano e affascinante. Non avevo allora una conoscenza approfondita sul genere del commentario antico, molti passaggi mi erano rimasti oscuri. Avevo però constatato l'interesse di un testo come quello e quanto poteva apportare alle nostre conoscenze sull'insegnamento letterario, sulla ricezione della storia del teatro, sulla grammatica etc. Avevo intenzione di allestire un seminario di traduzione // (p. 22) giacché tale testo era stato molto letto e utilizzato da parte dei dotti seicenteschi (La Mesnardière, Corneille, D'Aubignac, considerando solo autori francesi) ma nessuno aveva pensato di tradurlo *in extenso* in lingua moderna. Questa fonte di conoscenze mi sembrava, senza l'aiuto di una traduzione, persa da parte delle comunità scientifiche contemporanee, quali gli esperti di teatro o di teorie linguistiche, che avrebbero avuto grande interesse a riscoprirlo.

È allora che Bruno Bureau è intervenuto per dire che il progetto era appassionante ma che la natura stessa del testo costringeva a pensarlo oggi come un'edizione elettronica. Mi ha spiegato per quale motivo, mostrandomi che la delimitazione dei lemmi e degli scolii, la presenza all'interno degli scolii di citazioni interne ed esterne, la presenza nei lemmi o nelle citazioni di abbreviazioni, di segmenti autonomi, del greco, tutto costituiva un intarsio la cui

strutturazione non poteva essere esplicitata chiaramente se non attraverso una scrittura XML ; e, anche se non mi ha convinto immediatamente, ho accettato l'idea e abbiamo organizzato il seminario HyperDonat.

Si trattò innanzitutto, per gli studenti del Master 2, del dottorato e per i giovani ricercatori, di un seminario di traduzione. Uno di noi inviava a tutti gli altri una parte del testo di Wessner e la traduzione che ne proponeva; successivamente, durante una sessione di quattro ore, ognuno muoveva osservazioni al lavoro e ne proponeva dei miglioramenti. Si generavano discussioni appassionanti che hanno convinto Bruno Bureau e me della necessità della traduzione: era indispensabile avere una comprensione precisa prima di lavorare a un'edizione.

Abbiamo constatato che Wessner, che non aveva effettuato una traduzione, proponeva nella sua edizione dei passaggi non interpretabili, che non avrebbe edito in quel modo se li avesse dovuti tradurre. Abbiamo anche convenuto che talvolta avevamo solo l'impressione di comprendere: quando la frase di Donato era limpida e il suo contenuto senza interesse, tautologico, abbiamo tentato di capire quale contenuto implicito la potesse rendere più interessante. Abbiamo cercato sistematicamente il valore dello scolio: cosa voleva dire il commentatore? Cosa commentava?

Nel commento ad *Ad.* 533, ad esempio, per commentare la forma *calleo* Donato scrive *callide scio*. Sotto l'apparente tautologia però egli in realtà compie un'osservazione stilistica: ciò che dice è che qui *calleo* si costruisce come *scio* che è transitivo (ciò non avviene quando significa *callidus sum*). Altro esempio: a *Ph.* 367.3 lo scolio dice **3** *Et IN VITA hoc est: cum uiueret et quod a uiuo fit*. Che interesse ha nel precisare che *in uita* significa *cum uiueret*? In realtà questo commento si può interpretare in due maniere. O Donato indica che l'espressione *in uita* può rimandare sia alla durata della vita che a ciò che si compie durante la // (p. 23) *vita*, o vuole far comprendere che si tratta in verità del padre della giovane figlia e dunque *cum uiueret* spiega che bisogna intendere *in uita eius* e non *in uita mea*. Si può anche pensare a un'altra spiegazione senza alcun rapporto con questa ma morfologica e ortografica: Donato potrebbe voler indicare che *in uita* consta di due parole (non si tratta dunque dell'aggettivo *inuita*) e che si tratta del *uita* che vuol dire 'vivere' (*cum uiueret*) e che deriva da *uiuus* (*quod a uiuo fit*). Questa spiegazione sarà allora puramente gratuita giacché il testo terenziano non presenta alcuna ambiguità a tal riguardo. In ogni caso l'interpretazione non si può privare all'occasione di un'annotazione.

Non avendo individuato la *quaestio* non conosciamo esattamente l'intenzione dello scoliasta e nel particolare abbiamo compreso che il rapporto tra il lemma e lo scolio non è necessariamente quello che si crede. Questi primi approcci ci hanno resi progressivamente sensibili alla problematica ecdotica.

Contemporaneamente, mentre conducevamo sulla carta e su Word questi abbozzi di traduzione di Donato, riflettevamo sull'aspetto Hyper-, l'aspetto *DH* del nostro dittico. Con l'aiuto della sezione digitale dell'ENS di Lione e del nostro contatto con Maud Ingarao, che è e rimane il nostro ingegnere dall'inizio del progetto, abbiamo osservato e messo a confronto diversi sistemi di trattamento delle informazioni e abbiamo optato in un primo momento per il sistema di scrittura e di edizione *Arcane*, creato a Montpellier da Eric-Olivier Lochard. Una richiesta di finanziamento che abbiamo rivolto alla regione Rhône-Alpes ci ha concesso (con nostro grande stupore) di ottenere una borsa di dottorato di tre anni e dei fondi per acquistare la licenza del software e finanziare qualche spostamento del progettista verso Lione.

L'assunzione della dottoranda Emmanuelle Raymond (attualmente maître de conférences all'Università di Angers) ci ha consentito di intraprendere effettivamente il trattamento dei dati del commentario di Donato e la sua trasformazione in *corpus* digitale.

2. Dalla carta stampata al sito web

I primi due anni sono stati complessi. Abbiamo continuato a lavorare a partire da un testo Word che il software *Arcane* doveva poi trasformare in XML; doveva però essere controllata la resa finale e, di base, il risultato non era soddisfacente, sia perché il software (che è un «software-proprietario») era molto rigido e ogni modifica *ad hoc* non poteva essere effettuata se non dal suo creatore, ma analogamente perché il nostro progetto non era maturo. Abbiamo cominciato ad apprendere la tecnica senza sapere con esattezza ciò che volevamo fare. // (p. 24)

Questa è stata la nostra prima lezione, appresa con l'esperienza e con un anno di lavoro perso: bisognava avere un'idea precisa del progetto prima di cominciare a realizzarlo. Finalmente, dopo numerose discussioni tra i padri fondatori (Bureau, Nicolas, Raymond, Ingarao), sapevamo cosa doveva essere HyperDonat:

*sul piano giuridico, un sito gratuito con licenza *creative commons*: all'occorrenza una licenza CC by-nc-sa [by = paternità; nc = non commerciale; sa = condivisibile senza variazioni, secondo la codifica dei contratti *creative commons*]; essa consente all'utilizzatore, oltre l'accesso gratuito, il diritto

1 – di riprodurre, distribuire e diffondere quest'opera al pubblico;

2 – di modificare quest'opera secondo le condizioni seguenti:

2.1. Paternità. - Bisogna citare il nome dell'autore originale nella maniera seguente (ma non una maniera che suggerisce che si approva o meno l'utilizzo dell'opera): *Hyperdonat - Collection d'éditions numériques de commentaires anciens avec traduction, commentaire et annotation critique*. Accesso in linea il [data] a <http://hyperdonat.tge-adonis.fr>.

2.2. Nessun uso commerciale. Non vi è il diritto di usare quest'opera per fini commerciali.

2.3. Condivisione analoga delle condizioni iniziali. Se quest'opera viene modificata, trasformata o adattata non vi è il diritto di distribuirla se non sotto un contratto identico a questo. A ogni riutilizzo o distribuzione di quest'opera bisogna far apparire chiaramente in maniera pubblica le condizioni contrattuali della sua messa a disposizione. La maniera più appropriata di indicarle è un link a questa pagina web: <http://creativecommons.org/licenses/by-nc-sa/2.0/fr/>. Ognuna di queste condizioni può essere eliminata se si ottiene l'autorizzazione di titolare dei diritti sull'opera. Niente in questo contratto diminuisce o restringe il diritto morale dell'autore o degli autori.»¹

* sul piano tecnico, HyperDonat doveva essere un *corpus* in xml-TEI (Text Encoding Initiative). Il consorzio TEI ha stabilito delle regole tecniche di marcatura molto precise (le *TEI Guidelines*), in- // (p. 25) terattive (gli utilizzatori possono chiedere la creazione di marcature *ad hoc* adattate a un progetto specifico), costantemente in corso di costruzione. La TEI ci è

¹Aspetti legali del nostro sito su <http://hyperdonat.huma-num.fr/editions/html/mentions-legales.html>.

sembrata essere la risorsa maggiormente suscettibile di evoluzione tecnica, conseguentemente la più sicura per l'organizzazione dei dati. Si tratta evidentemente di una scommessa per il futuro

* sul piano della visualizzazione html, abbiamo creato una «sceneggiatura dell'utilizzatore»: ci siamo chiesti cosa avremmo voluto trovare sul sito se fossimo stati dei lettori occasionali, dei lettori interessati, dei lettori esperti. Abbiamo così costruito l'architettura del sito in funzione di questi possibili percorsi. La nostra sceneggiatura si è basata evidentemente sulla nostra pratica di utilizzatori di altri siti. Mi capitava ad esempio spesso di consultare l'eccellente sito CGL che riporta on-line il *corpus* integrale dei grammatici di Keil, realizzato sotto la direzione di A. Garcea e di V. Lomanto e che ha sede a Parigi ¹.

Qual è il vantaggio di un sito del genere? La sua funzione principale è l'accesso gratuito a un *corpus* voluminoso che consente di evitare di andare a consultare in biblioteca gli otto volumi di Keil. Non è però solo questo: esso presenta il testo secondo un aspetto filologico, segnalando i cambiamenti di pagina e di linea conformi all'edizione di riferimento (Keil o edizioni più recenti quando ve ne sono). L'utilizzatore può dunque continuare a citare questo passaggio di Cosenzio *prima positio diuiditur in nomen proprium et appellatiuum et epithetum* secondo una maniera che ogni utilizzatore riconoscerà dei volumi di Keil: Consentius, *GL* 5, 339, 1-2. Nessuna discontinuità nell'uso. Il testo è presentato in forma continua con il simbolo | per indicare i passaggi di linea e un numero tra parentesi quadre per segnalare l'inizio di una nuova pagina.

Vi è un valore aggiuntivo: i testi sono suddivisi per sezione (*de nomine, de aduerbio*, per esempio), anche quando questa suddivisione non è proposta nell'edizione cartacea di partenza. Ulteriore vantaggio che ne deriva: il sito presenta una sezione «Ricerca» che consente di fare una ricerca lessicale su una stringa di caratteri. Se si digita, pertanto, nel campo di ricerca «appellatiuum» si troverà, tra gli altri, la frase di Cosenzio prima citata ma anche i 106 altri passaggi del database che riportano la sequenza in questione con una linea di contesto e il riferimento a Keil. Si tratta dunque di un *index uerborum* meccanico. Coloro che hanno familiarità con la ricerca automatica possono anche servirsi // (p. 26) istintivamente degli operatori booleani: così la query «appellatiu*» integrerà alla richiesta le forme *appellatiua, appellatiue, appellatiuos*, etc. e proporrà 355 occorrenze.

La sceneggiatura dell'utilizzatore non è però ottimizzata. Il limite principale del sito è che non fornisce la possibilità di navigazione diretta tra le occorrenze trovate dal motore di ricerca e i testi tra i quali le ha ricercate. È necessario dunque navigare manualmente dalla sezione «Ricerca» a quella «Testi», scorrere la lista degli autori fino a Cosenzio, ad esempio, poi aprire la sezione interessata, poi trovare ciò che corrisponde alla linea 7 della pagina 412; poi ripetere l'operazione 354 volte. Un altro svantaggio secondo noi: il testo latino è essenziale, senza traduzione, senza apparato critico, senza note; è così destinato al lettore specialista che già si serviva di Keil. E mentre le numerose citazioni degli autori riportate dai grammatici latini sono, nel sito, marcate da un colore convenzionale, il loro riferimento non è riportato laddove nell'edizione cartacea è in una sezione infratestuale: in Keil dunque si può apprendere che si tratta di quello specifico passaggio virgiliano; questa informazione, nell'attuale versione del

²Cf. kaali.linguist.jussieu.fr/CGL/index.jsp

sito, non è invece reperibile. Il sito dunque non consente di fare completamente a meno di consultare il libro sul quale si basa.

Inoltre la volontà, nel CGL, di creare un fac-simile xml della pagina di Keil senza fotografarla è, nei dettagli, leggermente imprecisa. Keil e gli editori più recenti usati nel CGL sono stati impaginati da una stampante; succede (quasi per ogni pagina) che delle parole risultino suddivise da una linea alla successiva, talvolta con un salto di pagina. Questo però non accade mai nel sito CGL: le parole non sono suddivise. Ne risulta pertanto in qualche caso un piccolo spostamento di linea dalla prima parola della linea, perfino del suo numero di pagina.

Ritengo che il rifiuto di dividere in sillabe, che impedisce così di seguire fino alla fine la logica di riproduzione dell'edizione cartacea, sia motivata dall'intento di poter trovare ciascuna parola (delimitata da uno spazio alla sua destra oppure da un segno di interpunzione) nella ricerca lessicale. Il motore che cerca «*appellatium*» non troverà «*appella-tium*» che non è la medesima sequenza. Per evitare questo inconveniente, gli ideatori del CGL hanno deciso di non suddividere le parole, quale sia la maniera in cui esse si presentano nell'edizione cartacea. Esisteva però una soluzione xml semplice che consisteva nel trattare il trattino non come un carattere ma come un segno di separazione delle parole. Come il segno |, simbolo del passaggio di linea, è un'aggiunta (non si trova evidentemente in Keil), non interrogabile (il motore di ricerca non lo trova)³, la divisione // (p. 27) delle parole e il salto delle pagine potevano essere trattati come dei segni xml non interrogabili (i segni <lb> e <pb> con l'attributo "hyphen"). Il motore di ricerca può essere facilmente programmato per non considerare i segni e interessarsi solo alle stringhe di testo propriamente dette. In questo caso un segmento xml come «*appella*<lb rend="hyphen"/>*tium*» sarà individuato dal motore di ricerca con una richiesta del tipo «*appellatium*» e la visualizzazione, più fedele all'edizione cartacea, può dare facilmente «*appella-|tium*» se è stata fatta una raccomandazione sistematica e automatica sul trattamento html dell'interruzione di linea e del trattino. Potremmo in questa maniera, grazie a dei semplici mezzi tecnici, sia garantire il rispetto della *mise en page* sia la possibilità di far trattare dal motore di ricerca come appartenenti alla medesima voce tutte le varianti tipografiche occasionali: *appel-latium*, *appella-tium*, *appellatium*, etc.

Abbiamo appreso dalla consultazione costante di questo sito (e di altri analogamente rilevanti, come gli *Itinera Electronica* di Lovanio per esempio) quanto volevamo vedere a tempo debito sul nostro sito.

Non eravamo intenzionati a rappresentare un fac-simile di ogni pagina dell'edizione Wessner. In effetti, da un punto di vista filologico, il riferimento a Wessner (come a Thilo-Hagen per Servio ad esempio) non si effettua in questo modo: Don. *ad Ad.* II, p.112, 20 (W) *sapientissime Terentius* etc. ma così: Don. *ad Ad.* 541, 2 *sapientissime Terentius* etc. Non è importante sapere che lo scolio in questione comincia alla linea 22 della pagina 112 del secondo tomo dell'edizione di Wessner ma che si tratta del secondo scolio al verso 541 degli *Adelphoe*.

Non c'è pertanto bisogno di indicare i passaggi di linea, i salti di pagina. Se si desidera mettere a confronto il nostro testo con quello di Wessner basta cercare la pagina dell'edizione cartacea in cui si trova il commento al verso 541 degli *Adelphoe*, che si trova dopo quello al verso 540 e prima di quello al verso 542...

³ Il simbolo di passaggio di linea è indicato in questo modo nel *code source* di ogni pagina del CGL: |. Il segno è incluso in un tag e il testo del «text» è interrogabile dal motore di ricerca proposto.

3. La struttura del sito Hyperdonat

Questo ci ha fornito la completa struttura del sito: un commentario lineare come quello di Donato a Terenzio si basa sul testo commentato. Lo si vede bene nelle edizioni manoscritte o a stampa nelle quali il commentario, scritto in caratteri più piccoli, in forma breve attorno al testo commentato, è posto al centro in caratteri più ampi. È il testo di Terenzio che regola l'impaginazione, è esso che regola anche la struttura del documento xml. // (p. 28)

Volevamo che fossero visualizzati il testo di Terenzio, la sua traduzione, il commento di Donato e la sua traduzione ma secondo una disposizione tale che i quattro testi potessero essere contemporaneamente affianco. Per assicurare questa 'parallelizzazione' l'unica soluzione possibile suggeritaci dal nostro ingegnere consisteva nello strutturare quattro documenti xml-TEI distinti, quello del testo originale (to), del testo tradotto (tt), del commento originale (co), del commento tradotto (ct) nella stessa maniera per tutti ove fosse possibile. Ogni commedia di Terenzio consta di un prologo, poi cinque atti, ognuno caratterizzato da più scene; in ogni scena dei versi numerati *in continuum* per tradizione a partire dal verso primo del prologo. Ter. *Ad.* 541 è un riferimento assoluto universalmente riconosciuto nel mondo degli antichisti. Questo doveva rimanere la nostra unità di base. Il testo di Terenzio (to) è dunque organizzato sequenziato per versi (è il tag <l> per 'linea'), separato da dei <lb> (linebreak) ognuno caratterizzato da un numero preciso. Questi versi sono raggruppati in <lg> (line groups) che corrispondono alle scene; le scene sono raggruppate in <div> (division) che corrispondono agli atti. Il prologo è una <div> dello stesso livello dei cinque atti. In quest'ottica, to e tt sono strutturati nello stesso modo. Il gruppo dei <div> (prologo e cinque atti) forma un <body>; è il corpo del testo.

Il commento (co e la sua traduzione ct), invece, ha lo stesso <body> di to e tt: prologo, atti, scene, versi, ognuno con un numero sequenziale. L'atto 1 costituisce il medesimo raggruppamento dei versi con gli stessi numeri in cc, co, tt e to. Questo consente la parallelizzazione. Laddove un verso è vuoto (non c'è commento al verso 527 degli *Adelphoe* per esempio) il suo posto è analogamente previsto; laddove compaiono più scolii per un unico verso (528.1, 528.2, 528.3 per esempio), nel segmento che corrisponde a <l n="528"> si introduce ogni scolio differenziandolo con un <milestone> (ad esempio <milestone n="1" type "scholie"/>).

Co, inoltre, ha una specificità: il commento comincia sempre con una *Praefatio*. Non ci troviamo ancora nel commento puntuale ma in una sezione più sintetica sulla commedia commentata, la sua collocazione nell'ambito dell'opera terenziana, la sua datazione, il suo riassunto etc. Questa sezione (di cui operiamo una <div>) è un <front>. Il <front> di ct ha le stesse prerogative di quello di co: saranno dunque in parallelo a tempo debito.

L'insieme <front> + <body> costituisce un <text> incluso in un doc <TEI> (per esempio co); i quattro <TEI> (co, ct, to, tt) sono raggruppati in un <teiCorpus> (per esempio DonAde per *Donati Adelphoe*).

La ricerca funziona bene su un segmento omogeneo, co ad esempio: vi è l'integralità del testo di Donato strutturato in atti, scene, versi di Terenzio. In // (p. 29) questo documento xml vi è tutto in forma continua. Si può aggiungere un'annotazione, modificare il testo, cambiare un numero di scolio o invertirne due, esplicitare un'abbreviazione, correggere un errore. Sul sito però, secondo le nostre aspettative, il testo è discontinuo: si ha visivamente sotto gli occhi il

verso 541 di Terenzio come deriva da to, la sua traduzione generata da tt, il commentario di Donato (co) e la sua traduzione (ct). Il verso 542 di Terenzio non comparirà che una volta terminato il blocco di scolii che riguardano il verso 541. L'omogeneità dell'insieme è assicurata dall'unità «541» giacché supponevamo che tale progetto di lettura fosse il più adatto per un commento di questo tipo.

Questa struttura a quattro parti, to, tt, co e ct, può funzionare con ogni commento lineare, anche se non organizzato per versi. Nel caso di un testo letterario in prosa, l'unità sarà senza dubbio costituita dal paragrafo dei filologi, inserito in un capitolo, esso stesso all'interno di un libro etc. (per esempio lo pseudo-Asconio su Cicerone, progetto in corso). Si tratta per analogia dello stesso modello. Esso funziona invece meno bene per quella tipologia di commenti più 'sfilacciati' come ad esempio il commento – realizzato per Hyperdonat – al secondo libro della *Consolatio* di Boezio di Guglielmo di Conches (che pone inoltre il problema del prosimetro nell'originale: l'unità è talvolta il verso, talvolta il paragrafo) o il commento sintetico e non lineare che realizza Olimpiodoro al *Gorgia* di Platone (progetto al debutto): in questo caso il commentatore non interviene spesso se non per evidenziare un riferimento e il commento si sviluppa alla stregua di una lezione.

In questo caso si pone il problema di sapere se bisogna mettere in parallelo il testo di Platone: talvolta a tre pagine del *Gorgia* corrispondono dieci di Olimpiodoro. La metodologia dovrà dunque essere indubbiamente adattata al tipo di commentario.

4. Gli strumenti di ricerca

Rispetto alle procedure di applicazione, volevamo uno strumento comodo, dinamico e molto efficiente. In primo luogo, cercavamo un'interattività che manca nel sito CGL: una volta generato un documento dinamico derivante da una ricerca, bisognava che fosse possibile tornare dal segmento individuato al testo anche con un semplice clic.

Abbiamo così aggiunto alla richiesta semplice, nella quale si digita il testo nella sezione prevista, una richiesta X-Query che richiede delle competenze particolari. Coloro che conoscono il linguaggio X-Query possono ad esempio individuare in una maniera molto precisa tutti i termini identificati come auto- // (p. 30) nimi e nei quali si trova la sequenza *re po*; se gli utilizzatori non hanno compiuto degli errori di sintassi la macchina li indirizzerà a *Ad. 656.1* dove si trova un autonomo *dicere posse*.

Volevamo ancora di più. Al posto di una ricerca lessicale semplice basata su una sequenza di caratteri, eventualmente migliorata grazie agli operatori booleani, volevamo che fosse possibile svolgere una ricerca morfologica. Chi ricerca tutte le occorrenze di *scelus* può cercare prima *scelus*, poi *sceleris*, poi *sceleri* (e si rende conto allora che trova ciò che ha trovato sotto *sceleris* e che ha anche trovato *sceleribus...*), etc. Se è più esperto, richiederà *scelus* e poi *sceler** che gli darà tutti i termini che cominciano per *sceler* e succeduti (o no) da almeno una lettera. Avrà dunque tutto il paradigma di *scelus* ma anche delle forme non cercate quali *sceleratum*.

Abbiamo così elaborato un lessico completo di Donato in cui tutte le forme di un paradigma presenti nel commentario sono raggruppate sotto una 'forma etichetta'. Se si clicca su *scelus* si scopre che le forme usate almeno una volta da Donato sono *scelera*, *scelere*, *sceleribus*, *scelerum*, *scelus*, ma anche due forme che non sarebbero mai state rinvenute con la ricerca lessicale *sceler**: *scelvs* (in un lemma terenziano: i lemmi sono visualizzati in maiuscoletto con

la grafia v per u) e anche s. (in un lemma terenziano abbreviato); al contrario, la ricerca morfologica non si interesserà delle forme di *sceleratus* né di *scelerosus*.

4.1. Il Lessico Donato

Questo è effettivamente il valore aggiunto che ha implicato un grande lavoro di controllo. Abbiamo utilizzato per questo fine il lemmatizzatore *Collatinus* del gruppo di Lovanio (che propone online *Itinera Electronica*). Abbiamo integrato il testo completo di Donato (il co di tutti i nostri TEICorpus, quello degli *Adelphoe*, del *Phormio* etc. ma anche di Evanzio e della *Vita Terenti*). Con la risorsa di lemmatizzazione e in pochi minuti *Collatinus* ha prodotto i risultati: le forme concrete del testo erano state disposte ordinatamente sotto una forma etichetta (*tulisse* così sotto *fero*). E, fortunatamente, la macchina ha respinto alla fine della lista le parole sconosciute del suo dizionario, cosa che ha permesso di trattarle individualmente. Sono servite molte ore inoltre a comprendere quanto era noto a *Collatinus* e, eventualmente, male attribuito: per esempio *nominatiue* (interpretato *nominati* + *-ue*) non è nel nostro corpus una forma di *nomino* ma un avverbio (sconosciuto da *Collatinus*); *utique* non è una forma di *uti* né di *utor* (*uti* + *que*) ma l'avverbio *utique*, etc. Tranne qualche eccezione che rimane sul sito, la ricerca morfologica è però ormai funzionante e facilita molto la ricerca di parole greche, facendo evitare la digitazione dei caratteri greci nella zona di ricerca. Inoltre le parole // (p. 31) greche scritte sia in latino che in greco sono raggruppate con tutte le loro forme sotto la forma etichetta latina:

sotto il 'type' *metaphora* si ritroveranno così i 'tokens' seguenti: μεταφοῤῥᾱ, μεταφορά, μεταφοράν, *metaphora*, *metaphoram*.

Le etichette si presentano in questo modo:

```
<item type="normal">
  <label> <term>cogitatio</term> onis</label>
  <gen>f</gen>
  <pos>nom</pos>
  <item type="occurrence">
    <label>cogitatione</label>
  </item>
  <item type="occurrence">
    <label>cogitationem</label>
  </item>
  <item type="occurrence">
    <label>cogitationis</label>
  </item>
</item>
```

Si comprende in questo modo: *cogitatio, onis* è la forma <label>, quella che si trova nel dizionario, di un <term> *cogitatio*, che è di genere femminile (<gen>f</gen>), che è un nome

(<pos>nom</pos> [pos = «part of speech»]); questo label si trova sotto le forme reali (<item type="occurrence">) *cogitatione, cogitationem, cogitationis*.

4. 2 Il *Thesaurus*

Desideravamo anche la possibilità di servirci di una ricerca per nozioni e, per fare ciò, abbiamo pensato alla creazione di quello che chiamiamo il *Thesaurus* di Donato. La nostra idea, in termini di sceneggiatura-utilizzatore, era la seguente. La nostra ambizione era che la persona che ricercava, interessata a una precisa nozione, potesse ottenere grazie al *Thesaurus* tutti gli scoli donatiani che rimandavano alla suddetta nozione. Così la richiesta «sipario» oppure «scenario» doveva rinviare a tutti i passaggi (poco numerosi ma sparpagliati nelle 800 pagine dell'edizione Wessner) in cui Donato menziona un elemento scenografico o, nello specifico, il sipario. Come prevedere però tutte le richieste // (p. 32) possibili? Non essendo questo fattibile, abbiamo chiesto di orientare la scelta del lettore strutturando le nozioni presenti nel corpus.

In un *index rerum* cartaceo il lettore rimane spesso deluso. La nozione di cui è alla ricerca è spesso organizzata in ordine alfabetico; se non c'è può presentare (nel migliore dei casi) un rinvio (*pedagogia*: vedi *scuola*); in caso contrario, bisogna cercare intuitivamente un sinonimo che potrebbe esserci (*didattica? Lezione? Insegnamento?*).

Abbiamo dunque costruito un'*ontologia* completa del mondo di Donato così come risulta dal suo commentario⁴. Questa strutturazione si effettua per inglobamenti in successione. C'è ad esempio, in Donato, una rappresentazione del mondo animale; in questo mondo animale ci sono dei pesci, degli uccelli, dei mammiferi, degli insetti. Tra i mammiferi, degli animali selvatici e degli animali addomesticati. E, nella misura in cui è possibile, è utile presentare le nozioni in un sistema di inclusione di iponimi negli iperonimi: ogni uccello è un animale, l'inverso non è vero. Allo stesso tempo, alcuni uccelli sono degli animali domestici oppure in ogni caso di allevamento: tale nozione di 'animale domestico' non deve dunque essere esclusiva dei mammiferi. Era dunque necessario stabilire tra le nozioni delle relazioni di filiazione e di correlazione (questa nozione è 'figlia' di quest'altra, queste nozioni sono invece correlate senza essere in un rapporto di parentela).

Abbiamo molto riflettuto su questa strutturazione con il supporto scientifico di altri ambiti e di specialisti dell'epistemologia. Sono state compiute scelte molto difficili da rivedere costantemente.

Questa è ad esempio, tra le altre, una delle nostre esitazioni nella riflessione. Nel momento di classificare le parti del discorso latino per richiamare la grammatica di Donato, abbiamo creato chiaramente una nozione 'avverbio'. Ci siamo poi detti che se qualcuno avesse voluto cercare la nozione 'avverbio' avremmo dovuto presentargli tutte le occorrenze degli avverbi (analogamente per i nomi, i verbi etc.). Abbiamo allora marcato tutti gli avverbi di Donato con un tag speciale che permetteva al motore di ricerca nozionistico di trovare tutti gli avverbi sotto la richiesta 'avverbio'. Tale procedura era poco logica: confondevamo il nome e la cosa, la nozione e le sue realizzazioni grammaticali, un errore simile alla confusione tra il segno linguistico e il suo referente extralinguistico, ad esempio. Chi è interessato alla nozione 'avverbio' non cerca le occorrenze di *diu* o di *leniter* ma quelle di *aduerbium* o di *aduerbialiter*

⁴ Un'*ontologia*, nel sistema delle conoscenze, può definirsi come una «formal specification of shared conceptualization» (T. R. Gruber 1993, *Towards Principles for the Design of Ontologies...*).

e tutti i passag- // (p. 33) gi in cui Donato dice qualcosa sulla categoria 'avverbio'. Abbiamo quindi annullato il lavoro intrapreso ma abbiamo considerato che, di base, poteva essere utile individuare una categoria morfologica come si può individuare un termine del lessico e abbiamo deciso di marcare ciascuno dei termini del 'Lessico Donato' con le proprie caratteristiche morfologiche : nome (maschile, femminile, neutro), verbo (deponente, transitivo etc.)... Se si prova ora a conoscere tutti gli avverbi usati da Donato lo si può fare attraverso una richiesta X-Query giacché tutti gli avverbi del lessico sono marcati <pos>adverbe</pos>. Ciò può facilitare le ricerche sintattiche del tipo 'treebanking' alle quali non ci siamo fermati nel nostro progetto⁵.

Il rischio (che non abbiamo completamente evitato) è di presentare un'eterogenea *Weltanschauung*: bisognava rappresentare l'epistemologia di un erudito del IV secolo anche dove essa risultasse erronea secondo i moderni. Come la balena che per Donato è un pesce (si veda *Eun.* 257.5) e non un mammifero. La tentazione di correggere secondo il nostro punto di vista moderno è grande in numerosi campi. L'obiettivo da raggiungere, molto interessante di base, è però limitarsi alle conoscenze e alle nomenclature dell'autore del testo trattato purché se ne abbia un'idea chiara, cosa che non si verifica sempre.

Dopo numerose esitazioni e dopo un anno di lavoro per idearlo e per realizzarlo tecnicamente, abbiamo elaborato una *Weltanschauung* di Donato che si presenta orientativamente in questo modo.

Il mondo di Donato è descritto secondo tre grandi iper-nozioni di partenza: Umanità, Contesto, Ambiente naturale. Mettiamoci nella testa di un utilizzatore del sito che è interessato ai cetacei e andiamo a prendere la balena nel nostro Thesaurus. A priori se c'è una balena in Donato essa deve abitare l'Ambiente naturale. Se apriamo quindi il menu della scheda 'Ambiente naturale' si presentano cinque iper-nozioni intermedie: 'luogo naturale', 'elementi', 'mondi naturali', 'fenomeno naturale', 'natura'. L'istinto ci conduce verso 'mondi naturali' che contiene 'viventi' e 'non viventi'. Apriamo la scheda dei 'viventi' che si articola in 'animali' e 'vegetali'. Gli 'animali' si suddividono in 'mammiferi', 'uccelli', 'pesci', 'insetti', 'rettili e anfibi'. In un primo momento il lettore moderno aprirà 'mammiferi' per poi rendersi conto che non trova la balena nella lista 'mammifero' – 'mammiferi selvatici' – 'erbivori' / 'carnivori'. Se però ritorna sui pesci e clicca sulla lettera L (come lessico) allora il primo nome di pesce che trova è *cete*, il termine neutro plurale (greco) che indica numerosi cetacei. È là che si trova la balena tanto ambita. Sarà così arrivato all'unico scolio in cui si parla di balena (*Eun.* 257.5) ricercando nel campo 'recherche plein // (p. 34) texte' la parola francese 'baleine' a condizione di aver posto fra i parametri di ricerca la sezione Commentario tradotto (ct) oppure la parola latina *cete* nel Lessico.

Prendiamo ora il caso di un lettore che cercherà tutti i riferimenti fatti nel corpus di Donato a un autore teatrale. Può indubbiamente ricercare nel lessico, lettera per lettera, tutti gli antroponomi che reinviano a uno scrittore di teatro. Può però analogamente cercare una sintesi più completa grazie all'impiego del Thesaurus.

Apriamo questa volta l'iper-nozione di primo livello 'Umanità' (è un riflesso istintivo, suppongo). Nel ricercare delle individualità si apre 'individuo', poi 'prosopografia'. Si hanno

⁵Cf. per esempio nlp.perseus.tufts.edu/syntax/treebank/. Il treebanking è una tipologia di analisi sintattica ramificata dell'enunciato in scrittura xml.

allora tutti gli antroponimi delle persone reali presenti nel corpus. Tra di essi vi sono degli autori teatrali greci e latini, ma non solo.

Si arriva a un elenco più preciso se si torna indietro a 'individuo' e poi aprendo la sezione 'attività' e poi 'produzioni sociali', poi 'letteratura', poi 'scrittori': ivi compaiono due liste affini, quella degli 'scrittori greci' e quella degli 'scrittori latini'. Vi sono però tutti gli scrittori e non soltanto gli autori teatrali.

Se, invece, dopo 'umanità' ha aperto la scheda 'società' piuttosto che 'individuo' dal momento che cerca qualche cosa che, teatro o letteratura, sia un'attività sociale, la scelta più allettante può essere 'produzioni sociali'. Non si troverà però il teatro in questa sezione. Torniamo dunque indietro (attraverso dei semplici clic) e apriamo 'vita quotidiana': è ugualmente un vicolo cieco. Ritorniamo indietro e apriamo 'relazioni sociali' supponendo che il teatro antico implichi delle relazioni sociali, persino culturali. Dopo aver ancora errato un po' tra i 'codici sociali' ci troveremo probabilmente tra le 'attività sociali' che propongono 'tempo libero' in cui si trova 'spettacolo', poi 'tipologia di spettacolo', poi 'teatro'; si aprirà allora 'personaggi di teatro', poi 'autori di teatro' e si capiterà sulla lista alfabetica completa degli autori il cui nome è citato almeno una volta all'interno del commentario a Terenzio.

Si può notare come Afranio (che è una nozione di questa ontologia) si trovi in numerose liste contemporaneamente. Si capiterà su di lui, con l'intermediazione del Thesaurus, facendo una delle richieste seguenti:

Humanité ☐ individu ☐ prosopographie ☐ Afranius

Humanité ☐ individu ☐ activité ☐ productions sociales ☐ littérature ☐ écrivains ☐ écrivains latins ☐ Afranius

Humanité ☐ société ☐ productions sociales ☐ littérature ☐ écrivains ☐ écrivains latins ☐ Afranius

Humanité ☐ société ☐ relations sociales ☐ activités sociales ☐ loisirs ☐ spectacle ☐ type de spectacle ☐ théâtre ☐ personnes du théâtre ☐ auteurs de théâtre ☐ Afranius

// (p. 35)

E si toccano così i due principali difetti della nostra concezione di Thesaurus.

1. Alcuni campi, talvolta molto ricchi (per esempio le nozioni 'grammatica' o 'retorica') si trovano in numerose branche della ramificazione: un 'individuo', la sua 'attività', in connessione alle 'produzioni sociali' che si trova sotto 'società' e questo genera una duplicazione della ramificazione; ciò accresce notevolmente la ricerca del Thesaurus che diventa, di conseguenza, troppo lento;

2. La ricerca di nozioni finisce col diventare comunque una ricerca lessicale. Alla fine della stringa della richiesta si finisce nella scheda L che apre tutte le parole connesse alla nozione. Il nostro Thesaurus non è in definitiva che un'organizzazione differente delle parole del nostro lessico. Non sono le nozioni che ricerchiamo effettivamente ma le parole che noi stessi abbiamo associato a tali nozioni attraverso reinvi agli elementi del lessico.

Un esempio è quello della nozione 'armi' che si trova sotto 'vita quotidiana' – 'oggetti':

```

-<item n="vie quotidienne">
  -<item n="objets">
    -<item n="arme">
      <lexique n="arcus" />
      <lexique n="arma" />
      <lexique n="armatura" />
      <lexique n="armo" />
      <lexique n="claua" />
      <lexique n="clipeus" />
      <lexique n="dolo" />
      <lexique n="ensis" />
      <lexique n="funda" />
      <lexique n="gladius" />
      <lexique n="hasta" />
      <lexique n="iaculus" />
      <lexique n="inermis" />
      <lexique n="lancea" />
      <lexique n="machaera" />
      <lexique n="pharetra" />
      <lexique n="protelum" />
      <lexique n="sagitta" />
      <lexique n="scutum" />
      <lexique n="telum" />
    </item> (...) // (p. 36)
  </item>
</item>

```

La ricerca sulla nozione 'armi' avviene attraverso la ricerca degli elementi (vale a dire delle 'forme etichetta') che conducono esse stesse verso la nozione 'armi' (<ptr target="armes">). Prendiamo ad esempio la scheda lessicale di *clipeus*:

```

<item type="normal">
  <label> <term>clipeus</term> i</label>
  <gen>m</gen>
  <pos>nom</pos>
  <ptr target="arme"/>
  <item type="occurrence">
    <label>clipeum</label>
  </item>
</item>

```

Nel ricercare la nozione 'arma' il Thesaurus interroga dunque il lessico e gli chiede di trovare tutte le etichette che contengono il <ptr target="arme"/>; tra essi, *clipeus*. A questo punto il motore di ricerca ne individua le occorrenze (qui soltanto *clipeum*) in tutto il corpus di Donato per rimandare agli scolii che contengono un'occorrenza di *clipeus* (nell'occorrenza *clipeum* in *Ph.* 681.3 si trova infatti all'interno di una citazione virgiliana). Ripeterà l'operazione con le occorrenze di *arcus*, *arma*, *armatura*, etc. che contengono tutti il <ptr target="arme"/>.

Ammettiamo che con un ambito specifico come il vocabolario delle armi non manca alcuna occorrenza giacché quando Donato rimanda a tale sfera egli cita il nome di un'arma oppure l'iperonimo *arma*, tutti elementi marcati con il <ptr target="arme"/>. In altre situazioni però il rapporto tra la nozione e il lessico è molto più complesso.

Si verifica ad esempio il fenomeno della polisemia: i termini tecnici possono incrociarsi poiché un termine rientra in più nomenclature. Gli si possono attribuire più <ptr target="x"/> ma il motore di ricerca non farà differenza. Se si creano due etichette distinte, *causa 1* e *causa 2*, il primo può rimandare alla nozione filosofica di 'causa', il secondo a quella giuridica di 'processo'; le occorrenze reali di queste due **etichette** non verranno distinte e saranno trattate tutte come delle occorrenze di due **etichette** congiunte. In maniera analoga avviene per il termine *scaena* che significa sia lo spazio in cui si muovono gli attori sia l'unità testuale, parte dell'atto. Attribuiamo dunque all'etichetta *scaena* due <ptr target>: <ptr target="scene-lieu"/> e <ptr target="scene-texte"/>. La ricerca non sarà dunque sufficientemente specializzata sul piano delle nozioni a meno che, nel corpus di Donato, per caso una ricorrenza come *causae* sia sempre utilizzata in senso filosofico e un'altra, *causam*, sempre in senso giuridico. // (p. 37) In questo caso si può attribuire l'occorrenza *causae* all'etichetta *causa 1*, *causam* all'etichetta *causa 2*, ma non è che una falsa soluzione poiché tale organizzazione, che risulta vera per pura coincidenza, non sarà aprioristicamente vera per il corpus serviano se dobbiamo considerarlo con la *Weltanschauung* del latino del IV secolo pagano. Si tratta di una verità congetturale e non strutturale. Abbiamo quindi alla fine scelto di non attribuire il <ptr target="x"/> ai termini, in particolar maniera i verbi, troppo polisemici. Il motore di ricerca dunque ignorerà gli scoli interessati dalla nozione.

Al contrario, quando il motore mira al lessico, cercherà più volte lo stesso scolio. Si consideri lo scolio seguente:

Ph. 35.1: AMICVS SVMMVS MEVS ET POPVLARIS quod in omnibus fere comoediis, in quibus perplexa argumenta sunt, fieri solet, id in hac quoque Terentius seruat, ut προτατικὸν πρόσωπον, id est personam extra argumentum, inducat, cui dum ob hoc ipsum, quod ueluti aliena a tota fabula est, res gesta narratur, discat populus textum ex continentia rerum sitque instructus ad cetera.

Il nome dell'argomento vi si trova due volte, sotto due occorrenze differenti, *argumenta* e *argumentum*, che reinviano all'etichetta *argumentum* che è dotata di un <ptr target="argument"/>; se si cerca la nozione 'argomento' questo scolio sarà dunque richiamato due volte dal motore di ricerca. Donato inoltre utilizza in questo scolio la parola *personam* e la parola *πρόσωπον*: entrambe sono caratterizzate dal <ptr target="personnage-de-comedie"/>: anche a proposito di questo criterio nozionale, dunque, lo stesso riferimento sarà richiamato due volte e la cosa si verifica spesso.

Si pone analogamente il problema della nozione che si esprime in maniera diffusa, senza che vi sia connesso un termine preciso e pre-assegnato alla suddetta nozione. Ad esempio a *Ph. 259.2* c'è, secondo noi, un'allusione implicita all'*a parte*. Dal momento che però non si trova nessuno dei termini che rimandano agli <ptr target="aparte"/> o <ptr target="voix-basse"/>, questo scolio sfuggirà a chi andrà alla ricerca di tutte le occorrenze della nozione '*a parte*'.

Il limite della nostra metodologia del Thesaurus è dunque che impieghi eccessivo tempo di risposta; che non si ottengano tutte le risposte desiderate; che si ottengano dei riferimenti in qualche caso eccessivi.

Ora sappiamo che bisogna pensare alla relazione testo/nozione secondo una prospettiva differente. Non è il lessico che dovrà predeterminare la designazione di una nozione bensì lo scolio. Se uno scolio tratta di 'alberi' e di 'metafora' (*Ad. 239.2 nam inchoatium uerbum est, μεταφορᾶ ab arbore siccante facta, quae multis succisa ictibus ferri tandem incipit in casum pendere ruinamque // (p. 38) minari*) indichiamo allora queste due nozioni alla base dello scolio. Chiediamo così al motore di ricerca di non individuare più il lessico ma le nozioni 'alberi' o 'metafora' (o qualsiasi altra nozione della nostra ramificazione). Troverà allora tutti gli scoli marcati con 'metafora', compresi quelli in cui non compare il termine tecnico per indicarla ; e non li troverà che un'unica volta giacché gli scoli saranno marcati un'unica volta come 'metafora' anche se vi compaiono congiuntamente i termini *metaphoram* e *transferuntur*, entrambi attualmente marcati dal <ptr target="métaphore"/>.

C'è dunque bisogno di un nuovo lavoro colossale che consiste nello stabilire per ogni scolio quali nozioni del nostro Thesaurus vi siano presenti e nel marcarle, secondo una procedura da stabilire fra di noi. Quando ciò sarà compiuto ritengo che il nostro Thesaurus sarà una risorsa unica e inedita nel mondo delle *DH*: un *index rerum* completo, dinamico ed efficiente.

5. La collazione dei manoscritti: inizi di una procedura

L'intento attuale di HyperDonat consiste nel creare un'altra risorsa inedita. Stiamo tentando di individuare una procedura meccanica di collazione dei manoscritti in xml TEI. Per questo fine è stata impiegata una parte del lavoro dei quattro studenti napoletani ospitati a Lione nell'ambito del progetto *Gli studi classici e la società dell'informazione*, Michele, Gaia, Sara e Federica. Grazie al loro lavoro pionieristico, alle loro domande e ai loro commenti, avanziamo speditamente in questa direzione grazie all'aiuto di una stagista, Ariane Pinche. Riparleremo senza dubbio di questa risorsa quando sarà meglio sviluppata e quando sapremo con esattezza cosa ne vorremo fare.

È tempo di concludere. Spero di avervi convinto che la tecnologia e il linguaggio xml sono un valore aggiunto per le nostre discipline. È fondamentale considerare che il lavoro che si può intraprendere in direzione delle *DH* non è un rinunciare alla filologia. Nessuno se non un filologo può decidere che quel termine del lessico è un sostantivo femminile; solo uno specialista della società del IV secolo può distinguere il mondo tardo-antico in categorie nozionistiche per rappresentarne una visione del mondo. Non si tratta di una semplice questione di tecnica: si tratta innanzitutto dell'epistemologia della disciplina.

È questa una risorsa che migliora le nostre performance, che velocizza l'accesso alle conoscenze, le organizza in una maniera dinamica ma che obbedisce, evidentemente, a un linguaggio formale molto complesso, la cui gestione è una competenza in sé. Non si tratta però, per un filologo, di cambiare discipli- // (p. 39) na e di dedicarsi alla programmazione informatica e al suo sviluppo, quanto piuttosto di collaborare a stretto contatto con l'ingegneria. Non bisogna accontentarsi, dall'alto della propria torre d'avorio, di un lavoro in formato word per poi affidarlo a un informatico. Bisogna lavorare insieme a un progetto comune. Bisogna far comprendere all'ingegnere ciò che si fa, ciò che si vuole ottenere dal prodotto on-line, quali risorse devono andare sviluppate; bisogna che l'ingegnere discuta questo punto di vista, faccia osservazioni a monte necessarie, poi che il lavoro si testi in corso d'opera, si modifichi in

funzione dell'esperienza, si migliori. Bisogna che, attraverso tale collaborazione, il filologo comprenda una parte del trattamento tecnico, proposto al suo turno di correzioni. Da questo dialogo produttivo nascono grandi risultati di cui può beneficiare l'intera comunità.

ABSTRACT

Here is described how was conceived the HuperDonat project, how, by the comparison with other sites involved in Digital Humanities our project grew rich and began to have increased performances. It is particularly dealt with the epistemological aspects of the creation of query tools such as our HyperDonat Lexicon and HyperDonatThesaurus.