

Copisti Digitali e Filologi Computazionali

Federico Boschetti

ROMA 2018

Copisti Digitali e Filologi Computazionali

Federico Boschetti

ROMA 2018

Federico Boschetti (federico.boschetti@ilc.cnr.it)
Copisti Digitali e Filologi Computazionali

© CNR Edizioni, 2018
P.le Aldo Moro 7
00185 Roma

www.edizioni.cnr.it

ISBN 978 88 8080 297 6 (edizione digitale)
ISBN 978 88 8080 296 9 (edizione a stampa)

Sommario

1	Introduzione	1
1.1	Filologia collaborativa e cooperativa	1
1.2	Copisti digitali e filologi computazionali	4
2	L’acquisizione del testo digitale	11
2.1	Stato dell’arte	12
2.2	Metodi per migliorare l’OCR	14
2.3	Risultati sperimentali	21
2.4	Il processo di correzione manuale	24
2.5	Rimappatura del testo sull’immagine della pagina	24
3	Analisi di apparati e repertori	31
3.1	Le linee guida della Text Encoding Initiative	36
3.2	Marcatore dei repertori di congetture	40
3.3	Elaborazione dei documenti marcati	44
4	Trattamento automatico di apparati e repertori	51
4.1	Stato dell’arte	51
4.2	Edizioni critiche di riferimento e repertori di congetture	53
4.3	Formalizzazione della struttura di apparati e congetture	55
4.4	Metodi di allineamento	59
5	Analisi sintattica	63
5.1	La linguistica dei <i>corpora</i>	74
5.2	Dall’interesse linguistico all’interesse filologico	89
5.3	Ancient Greek Dependency Treebank	90

SOMMARIO

6	Modelli e risorse lessico-semantici	93
6.1	La semantica strutturale	94
6.2	WordNet(s)	120
6.3	Ancient Greek WordNet	144
7	Esplorazione degli spazi semantici	145
7.1	Struttura, costruzione e proprietà degli spazi semantici	146
7.2	Osservazioni sui cambiamenti diacronici di significato	150
7.3	Relazioni semantiche osservate	153
7.4	Spazi semantici e generi letterari	164
8	Commento a tre luoghi scelti	171
8.1	<i>Pers.</i> 114–139 e 532–547	171
8.2	<i>Pers.</i> 598–602	196
8.3	<i>Pers.</i> 704	202
9	Conclusioni	207
	Bibliografia	213

1

Introduzione

Il presente lavoro rielabora e integra materiali scritti per la tesi di dottorato in filologia classica sostenuta nel 2005 presso l'Università di Trento (in cotutela con l'Università di Lille III), dal titolo: *Saggio di analisi linguistiche e stilistiche condotte con l'ausilio dell'elaboratore elettronico sui Persiani di Eschilo*, e materiali scritti per la tesi di dottorato in *Cognitive and Brain Sciences: Language, Interaction and Computation* nel 2010 presso l'Università di Trento, dal titolo: *A Corpus-Based Approach to Philological Issues*.

Come nei due lavori originari, da un lato vengono descritti modelli di rappresentazione del testo, della variantistica e delle informazioni linguistiche e stilistiche che insistono sul testo digitale; dall'altro vengono illustrati metodi computazionali e statistici che permettono di acquisire il testo tramite il riconoscimento ottico dei caratteri (OCR) ed estrarre o elaborare nuove informazioni a partire da *corpora* digitali annotati linguisticamente.

1.1 Filologia collaborativa e cooperativa

Data la natura interdisciplinare di questo studio, sembra opportuno cominciare proponendo alcune definizioni. Nel decennio passato, convegni e conferenze dedicati a *filologia digitale* (CIULA E STELLA, 2007), *e-filologia* (ZURLI E MASTANDREA, 2009) e *filologia computazionale* (BOZZI, 2004a) facevano intendere che nella letteratura scientifica questi termini fossero usati quasi come sinonimi.

1. Introduzione

Grazie al crescente interesse nazionale e internazionale per le *Digital Humanities* (TERRAS ET AL., 2016), molte discipline umanistiche tradizionali hanno visto nascere o crescere la loro controparte digitale (ad esempio: archeologia digitale o filosofia digitale). Per questo motivo, il termine *filologia digitale* attualmente tende a prevalere sulle altre diciture per indicare qualsiasi interazione fra l'informatica e le discipline filologiche.

Tuttavia è utile individuare tre aree distinte, nelle quali sono richieste competenze diverse. Si propone quindi di destinare il termine *e-filologia*, per analogia con *e-learning*, alla progettazione, all'implementazione e alla manutenzione dell'infrastruttura telematica che consente la condivisione di strumenti e risorse tra gli studiosi (BABEU, 2011; CRANE ET AL., 2009). Il termine *filologia computazionale*, per analogia con *linguistica computazionale*, dovrebbe riguardare invece lo sviluppo di metodi e strumenti per analizzare linguisticamente e stilisticamente *corpora* testuali, per estrarre automaticamente informazioni dalla letteratura secondaria, etc. Infine, il termine *filologia digitale*, in una accezione più restrittiva di quella vista al paragrafo precedente, dovrebbe riguardare la digitalizzazione di edizioni a stampa, la creazione di edizioni diplomatiche e di edizioni critiche digitali, etc.

In linea con l'*Open Philology Project* dell'Università di Lipsia (FRANZINI ET AL., 2013), si preferisce usare come termine generale *filologia collaborativa e cooperativa* (DEL GRATTA ET AL., 2016), per spostare l'attenzione dai mezzi (modelli e risorse digitali, metodi e strumenti computazionali, infrastrutture telematiche) ai fini (dati aperti, modelli e metodi condivisi, risorse interoperabili e strumenti riusabili), secondo criteri di riproducibilità e di falsificabilità dei risultati, propri del metodo scientifico.

Anche se nel linguaggio comune *collaborare* e *cooperare* sono spesso usati come sinonimi, qui si adotta per analogia la distinzione proposta in ambito pedagogico fra *collaborative* e *cooperative learning* (KOZAR, 2010). La collaborazione implica continue interazioni e negoziazioni fra le parti in causa, con lo scopo di produrre un risultato finale in sé concluso. Ad esempio, l'edizione critica digitale di un'opera può essere frutto della collaborazione di più studiosi. La cooperazione implica

1.1 Filologia collaborativa e cooperativa

che il lavoro sia eseguito dalle parti in causa produca risultati (anche parziali) tali da poter essere riassemblati da terze parti in modi non previsti inizialmente. Ad esempio, lemmatizzatori e strumenti per l'analisi metrica, nati inizialmente in progetti indipendenti, possono essere messi insieme per eseguire l'analisi delle formule omeriche.

L'applicazione dell'informatica al testo nasce a metà del Novecento per affrontare temi di natura filologico-interpretativa, producendo indici e concordanze digitali (PASSAROTTI, 2013). Fino alla fine del secolo scorso era facile confondere la filologia digitale e computazionale con la linguistica computazionale (vedere DEGANI, 1992, per una critica a questo approccio), a causa dell'applicazione degli stessi metodi e delle stesse tecniche, senza specifici adattamenti. Ma dal punto di vista della linguistica computazionale, i testi sono sequenze seriali di unità testuali, mentre dal punto di vista della filologia digitale e computazionale, i testi, con la relativa variantistica, sono sequenze parallele di unità testuali che insistono sulle medesime posizioni testuali. Ai tempi delle osservazioni di Degani, trascurando le esigenze specifiche degli studi filologici, la creazione di indici e concordanze digitali era per lo più basata su singole edizioni senza apparato critico. Tuttavia, se non c'è modo di distinguere tra le lezioni attestate dai manoscritti e le lezioni congetturali, è impossibile valutare se i fenomeni linguistici e stilistici sotto osservazione sono genuini, probabili corruzioni del testo per mano dei copisti o brillanti soluzioni proposte da studiosi moderni, senza il supporto della tradizione.

BOZZI (2004b) sottolinea il passaggio dai metodi "artigianali" della filologia tradizionale ai processi "industriali" promossi dalla filologia digitale e computazionale, dove i testi sono la materia prima, elaborata da gruppi di specialisti, al fine di creare i prodotti finali: dizionari, indici, concordanze, etc. Ogni forma di razionalizzazione industriale richiede un compromesso tra la perdita dei dettagli e il guadagno in termini di scalabilità e oggettività. L'attenzione si sposta quindi dal prodotto finale (edizione come libro statico) al processo (edizione come aggregato dinamico di funzionalità che possono essere migliorate).

1.2 Copisti digitali e filologi computazionali

Le biblioteche digitali possono crescere in due direzioni: ampiezza e profondità. Nel primo caso, la biblioteca si estende grazie all'acquisizione di nuove opere digitalizzate. Nel secondo, la biblioteca si arricchisce di molteplici edizioni digitali e studi della medesima opera.

I *corpora* di testi greci e latini più completi, come il *Thesaurus Linguae Graecae* (TLG) e la raccolta di testi latini del Packard Humanities Institute (PHI), sono basati sulle più autorevoli edizioni critiche di ciascun autore classico. In queste raccolte, solo il testo stabilito dall'editore è stato digitalizzato, mentre l'apparato critico è stato omesso. Tali collezioni, anche se utili per una prima consultazione, non sono tuttavia adatte a studi filologici approfonditi, perché obbligano lo studioso a verificare continuamente sulle edizioni critiche a stampa la natura delle informazioni trovate. Il filologo deve identificare le varianti del manoscritto e le congetture degli studiosi, al fine di valutare quale sia la lezione più probabile, accettando o rifiutando le ipotesi degli editori precedenti. Inoltre il filologo ha bisogno di esaminare commenti, articoli e monografie riguardanti parti specifiche del testo. Pertanto, l'estensione in ampiezza delle suddette collezioni deve essere integrata dall'estensione in profondità, secondo i criteri della nuova generazione di biblioteche digitali (CRANE ET AL., 2006; STEWART ET AL., 2007).

Dovendo affrontare il testo in profondità, gli studi filologici sono necessariamente focalizzati su singoli autori, generi o periodi, anche se hanno bisogno di trovare collegamenti e paralleli nell'intera letteratura greca e latina. Per questo motivo, i gruppi di ricerca formati da specialisti devono condividere un'infrastruttura comune, come sottolineato da CRANE ET AL. (2009).

Si vede quindi che progetti e iniziative indipendenti tendono sempre più a convergere nei metodi e a mettere a disposizione della comunità modelli, risorse e strumenti. Il *Perseus Project*¹ continua ad ampliare l'infrastruttura per connettere diversi progetti riguardanti le scienze dell'antichità. Il progetto *Musisque*

¹<http://www.perseus.tufts.edu>. Tutti i link sono stati verificati il 15 gennaio 2018

1.2 Copisti digitali e filologi computazionali

*Deoque*² ha creato una piattaforma web per gestire varianti testuali di testi latini. Il progetto *Multitext Homer*³, anche se si concentra su un singolo autore antico, ha sviluppato modelli di citazione che possono essere estesi ad altri autori.

Alludendo al titolo dell'opera di REYNOLDS E WILSON (1991), anche nell'era digitale possiamo identificare copisti e filologi, per quanto la distinzione fra i due gruppi non sia e non debba essere netta.

Come evidenziato da RONCAGLIA (2010), gli ultimi decenni sono interessati da un mutamento epocale nella trasmissione dei testi. E come è accaduto dall'oralità alla scrittura, dal papiro al *volumen*, dal manoscritto all'edizione a stampa, anche in questo ultimo passaggio dal libro all'edizione digitale il cambiamento di supporto implica il ripensamento delle strutture testuali.

La digitalizzazione dei classici può essere suddivisa in varie fasi. Il primo passo è stato il completamento di corpora basati sui testi di edizioni canoniche spogliate dell'apparato critico e di tutte le informazioni paratestuali come prefazione, introduzione e indici. Uno dei prodotti più rappresentativi di questa fase è il *Thesaurus Linguae Graecae*, digitalizzato manualmente da operatori privi di competenze specifiche in filologia classica.

Il secondo passo è la digitalizzazione manuale delle varianti. Questa fase richiede sia la competenza del filologo, che deve identificare i criteri di selezione, sia la competenza dell'informatico, che deve creare strumenti adatti a gestire le varianti. Un apparato critico è una selezione, non un repertorio esaustivo di varianti e congetture. Il progetto *Musisque Deoque*, illustrato in MASTANDREA (2009), ha dotato di apparati digitali minimi una grande quantità di testi poetici latini. MONDIN (2009), collaboratore del progetto, osserva come nella stesura dell'apparato critico si possano stabilire criteri differenti a seconda degli scopi: per uno studio linguistico varianti ortografiche e morfologiche non possono essere omesse, ma per lo studio dell'intertestualità, in un apparato minimo hanno la priorità varianti lessicali e semantiche rispetto a varianti di altro tipo.

²<http://www.mqdq.it>

³http://chs.harvard.edu/chs/homer_multitext

1. Introduzione

La terza fase della digitalizzazione segue due linee: da un lato la digitalizzazione delle edizioni critiche a stampa e dall'altro la digitalizzazione della letteratura secondaria (commenti, articoli, monografie) e terziaria (dizionari, enciclopedie, etc.). Il ricorso a tecniche che aumentino l'accuratezza dell'OCR, l'estrazione automatica di informazioni, la mappatura della variantistica sulle edizioni di riferimento e il collegamento fra risorse diverse sono necessari per gestire la massa di informazioni.

Come si è visto, non è necessario che il copista digitale in senso stretto, che acquisisce il testo copiandolo manualmente o correggendo l'OCR, abbia specifiche competenze linguistiche o informatiche. Il copista-filologo deve invece essere uno specialista del testo che sta digitalizzando e, qualora non faccia uso di interfacce di inserimento dei dati create da altri, deve possedere nozioni sulla codifica dei testi.

La filologia computazionale si occupa dell'allineamento di varianti, dell'estrazione automatica di informazioni da apparati critici e dalla letteratura secondaria, dell'analisi linguistica e stilistica condotta con metodi stocastici, etc., operando sia su corpora sincronici (ad esempio il *corpus* della tragedia attica) che diacronici (ad esempio il *corpus* dei testi filosofici dall'età arcaica all'età del tardo impero). Secondo HILPERT E GRIES (2009), “the use of corpora that are divided into temporally ordered stages, so-called *diachronic corpora*, is becoming increasingly wide-spread in historical corpus linguistics, creating a natural bridge between corpus linguistics and computational philology”.

La filologia computazionale mutua metodi, strumenti e risorse dalla linguistica computazionale e, più in particolare, dalla *corpus linguistics*. I filologi computazionali devono quindi avere competenze sia nell'ambito filologico, sia nell'ambito informatico.

1.2 Copisti digitali e filologi computazionali

In questo lavoro, la scelta dei *Persiani* come caso di studio principale è motivata dalla ricerca di un equilibrio fra la complessità del testo poetico eschileo e la relativa semplicità dei livelli interpretativi rispetto ad altre tragedie più tarde di questo autore. Il contenuto non mitologico dei *Persiani*, organizzato intorno a sequenze di nomi di persona, di popoli, di città e le lunghe sequenze descrittive di luoghi e azioni rende più agevole l'analisi sul piano dei referenti e di conseguenza anche l'interpretazione sintattica ne è avvantaggiata.

Il lavoro, iniziato nel 2000, è stato possibile grazie alla cortese disponibilità del C.I.P.L.-L.A.S.L.A. di Liegi, che ha fornito il *corpus* digitale delle tragedie di Eschilo⁴ come tabella di un database relazionale contenente il riferimento al verso e alla parola del testo, la forma, il lemma, la categoria grammaticale di ciascuna parola, la *dramatis persona*, la punteggiatura e alcune altre informazioni, come ad esempio la corrispondenza fra fine di frase e fine di verso. Il *corpus* riproduce MURRAY (1955), cioè la seconda edizione di Murray, tuttavia in una tabella a parte sono contenute varianti e congetture accolte nelle successive edizioni di PAGE (1972) e di WEST (1998).

Attualmente il *Perseus Project* mette a disposizione online⁵ i testi delle tragedie di Eschilo, fra cui i *Persiani*, con le eccellenti analisi morfosintattiche condotte da Francesco Mambrini nel 2009.

Il presente studio è suddiviso in nove capitoli che seguono il naturale flusso di lavoro dall'acquisizione del testo fino alla sua analisi.

Questo primo capitolo ha tentato di illustrare molto brevemente il contesto in cui è nato e si è sviluppato lo studio e gli obiettivi, nonché i limiti, che si è posto.

Il secondo capitolo è dedicato all'acquisizione del testo digitale. L'OCR può essere applicato alle edizioni critiche del XIX e XX secolo, raggiungendo fino al 99% di accuratezza sul testo e oltre il 90% sull'apparato critico⁶. Queste prestazioni sono ottenute grazie all'allineamento e alla fusione di tre diversi output

⁴L'*index verborum* e le liste di frequenza prodotte tramite l'elaborazione dei dati del *corpus* digitale sono pubblicati in RIGO (1999).

⁵http://perseusdl.github.io/treebank_data

⁶In ogni caso, è importante sottolineare che l'apparato critico è in media inferiore al 10% della pagina.

1. Introduzione

dell'OCR e all'applicazione di un sistema automatizzato di controllo ortografico. Dopo un addestramento adeguato, tre sistemi di OCR sono in grado di gestire il Greco politonico. Ogni sistema di OCR è più o meno affidabile per caratteri specifici, ma il sistema di fusione sviluppato in questo lavoro calcola il carattere più probabile in ogni posizione e il risultato supera significativamente le prestazioni dei singoli sistemi.

Il terzo capitolo è dedicato all'analisi della struttura di apparati critici e repertori di congetture basati sul testo dei *Persiani* di Eschilo. I repertori di congetture non solo registrano le correzioni al testo suggerite dagli editori nelle proprie edizioni critiche, ma anche le proposte di emendamento contenute in commentari e articoli.

Il quarto capitolo è dedicato al trattamento automatico di apparati e repertori, in particolare per effettuare l'allineamento di varianti e congetture sul testo dell'edizione di riferimento. I repertori di congetture hanno una struttura ricorrente: infatti, oltre il 90% delle singole note sono costituite dal riferimento al verso interessato, dal testo della congettura e dal nome dello studioso che ha fatto la proposta. Un *parser* identifica questi blocchi d'informazione e viene applicato un algoritmo di allineamento per trovare la posizione esatta nel verso in cui le congetture devono essere collocate.

Il quinto capitolo è dedicato all'analisi sintattica. Vengono presi in considerazione diversi modelli (generativo-trasformazionale, lessico-funzionale, a dipendenze) e si cerca di spiegare il valore delle basi di dati sintattici (*treebanks*) attualmente a disposizione, mostrando come interrogarle e quale tipo di ricerche è possibile fare quando si hanno a disposizione informazioni sintattiche organizzate in modo coerente e sistematico.

Il sesto capitolo discute la creazione di risorse lessico-semantiche per il Greco antico. Dal punto di vista teorico, viene illustrata la struttura di una rete lessico-semantica modellata su *Princeton WordNet* e si mettono in vista i limiti dell'adattamento alle lingue antiche di modelli pensati per le lingue moderne. Dal punto di vista pratico, vengono spiegate le tecniche per estrarre informazioni da dizionari bilingui e proiettarle su una rete lessico-semantica già esistente.

1.2 Copisti digitali e filologi computazionali

Il settimo capitolo è dedicato all'esplorazione degli spazi semantici del Greco antico. Nella prima parte si spiegano i principi teorici e le modalità di costruzione degli spazi vettoriali su cui gli spazi semantici sono basati. Nella seconda parte invece si passano in rassegna i risultati ottenuti applicando il metodo esplorativo a diversi *subcorpora* ripartiti per genere (epico, tragico, filosofico) e si tenta di mostrare come le caratteristiche di tali generi emergano dalle associazioni semantiche generate dalle statistiche applicate ai testi.

L'ottavo capitolo è uno studio su luoghi scelti dei *Persiani* di Eschilo che parte dal metodo tradizionale del commento lineare basato sul vaglio critico delle fonti primarie e della letteratura secondaria, per aprirsi ad osservazioni rese possibili dalle risorse e dagli strumenti offerti dalla filologia collaborativa e cooperativa.

Il nono capitolo, in conclusione, discute come la filologia collaborativa e cooperativa possa supportare gli studi di filologia classica e riassume i risultati raggiunti.

2

L'acquisizione del testo digitale

Questo capitolo descrive parte del *work-flow* ideato presso il *Perseus Project*, in collaborazione con università e istituti nordamericani ed europei¹, per popolare una biblioteca di edizioni critiche a partire da testi greci e latini digitalizzati tramite OCR².

Il *work-flow* completo prevede che i testi vengano catalogati³, che le immagini delle pagine vengano acquisite online⁴ oppure tramite scanner, che venga applicato l'OCR alle immagini ottimizzate, che vengano applicati metodi computazionali per il miglioramento dell'accuratezza dell'OCR, che i risultati siano corretti manualmente e infine che i testi corretti siano rimappati sulle immagini originali.

Come già accennato nell'Introduzione, le prime grandi iniziative di digitalizzazione dei classici antichi scartavano o ignoravano le informazioni paratestuali costituite da prefazioni, introduzioni, indici, bibliografie, note e apparati critici. Ma i nuovi *corpora* greci e latini devono essere concepiti come collezioni di edi-

¹Per i dettagli sui progetti cui ha partecipato l'autore si vedano BOSCHETTI ET AL. (2009) e BOSCHETTI (2014).

²Optical Character Recognition

³Il catalogo dei testi è consultabile online all'indirizzo: <http://catalog.perseus.org>

⁴Fra le iniziative di digitalizzazione delle immagini delle pagine di libri a stampa si possono menzionare Internet Archive (<http://archive.org>), ad accesso libero, e HathiTrust Digital Library <https://www.hathitrust.org>, ad accesso limitato sulle immagini di alta qualità per i membri consorziati

2. L'acquisizione del testo digitale

zioni digitali e non come raccolte di semplici testi digitalizzati. Visto che il testo antico che si legge nelle edizioni moderne è il prodotto delle scelte dell'editore, che accoglie le varianti più probabili e le migliori congetture formulate dagli studiosi precedenti, il filologo ha quindi la necessità di avere a disposizione sia il testo che le informazioni paratestuali.

Tuttavia, le edizioni critiche dei classici sono difficili da riconoscere da parte dei sistemi di OCR per molteplici ragioni. Il layout della pagina è diviso in più flussi di testo con differenti dimensioni di carattere: il testo dell'opera come stabilito dall'editore, l'apparato critico e, facoltativamente, note storico-letterarie o traduzione a fronte. Poi, il Greco antico fa uso di un set di caratteri molto esteso per rappresentare le combinazioni di spiriti e accenti. Inoltre, le edizioni critiche sono tipicamente poliglotte, perché l'apparato usualmente è in Latino, i nomi degli studiosi citati sono in Inglese, Tedesco, Francese, Italiano o altre lingue moderne e la prefazione, l'introduzione, l'eventuale traduzione e gli indici sono parimenti in Latino o in una lingua moderna. Infine, le edizioni critiche dei secoli scorsi, di grande interesse per il lavoro del filologo, possono presentare pagine danneggiate.

La parte del *work-flow* di cui ci si occupa nel seguito riguarda le tecniche per migliorare l'accuratezza dell'OCR in post-processing⁵ e la creazione di strumenti di supporto alla correzione manuale collaborativa.

2.1 Stato dell'arte

L'attività condotta presso il *Perseus Project* sull'OCR applicato al Greco antico ha preso le mosse dalle linee guida espresse in STEWART ET AL. (2007). In quella ricerca, il riconoscimento di spiriti e accenti del Greco nelle edizioni a stampa non era stato preso in considerazione a causa dei limiti tecnologici imposti dai sistemi di OCR disponibili all'epoca dello studio.

Il primo passo per ottimizzare l'OCR è il *pre-processing* delle immagini che contengono testo (GUPTA ET AL., 2007). Attualmente esistono applicazioni, co-

⁵Si veda PIOTROWSKI (2012).

me *ScanTailor*⁶ e librerie grafiche, come *OpenCV*⁷ che facilitano il trattamento delle immagini ai fini dell'OCR. Se le immagini sono stati acquisite tramite scanner, si deve infatti procedere al riconoscimento e alla divisione della pagina di sinistra e della pagina di destra (*splitting*), alla correzione dell'eventuale rotazione (*deskewing*) della pagina, al riconoscimento degli specchi di testo sulla pagina (*content box recognition*), all'eliminazione di macchie (*despeckling*), all'eventuale stiramento dello specchio di testo (*dewarping*) e infine alla binarizzazione (*binarization*) con i valori ottimali di luminosità e contrasto. Per l'uso di diversi livelli di binarizzazione ai fini del miglioramento dell'OCR si veda LUND ET AL. (2013) e per il riconoscimento ottimale degli specchi di testo in casi difficili, con contenuti disposti su più colonne, si veda ROBERTSON ET AL. (2014).

Nel caso delle edizioni critiche, è utile il metodo proposto da RINGLSTETTER ET AL. (2005) per distinguere i diversi tipi di carattere in testi poliglotti, perché permette di trattare in modo diverso il testo greco dai testi (o dai brevi frammenti d'informazione) in Latino o in lingue moderne.

L'allineamento di molteplici output dell'OCR, che costituisce il centro del lavoro illustrato in questo capitolo, è utile non solo per migliorare l'accuratezza dei sistemi di riconoscimento presi singolarmente, ma anche per valutare i risultati, grazie al confronto automatico fra ciò che si è ottenuto in automatico e una trascrizione eseguita manualmente. CECOTTI E BELAÏD (2005) e LUND E RINGGER (2009), dopo aver allineato molteplici output dell'OCR, hanno applicato efficaci strategie di selezione fra le parole o fra i singoli caratteri allineati. YALNIZ E MANMATHA (2011) usano l'allineamento per scopi di valutazione dei risultati ottenuti.

Contribuiscono al miglioramento dell'accuratezza anche l'uso di risorse linguistiche e metrico-stilistiche e l'estrazione di informazione dai testi stessi, anche se riconosciuti in maniera imperfetta. Ad esempio, BOZZI (2000) ha illustrato una tecnica integrata fra OCR e correzione ortografica automatica (spell-checking) applicata a documenti testuali danneggiati. NAMBOODIRI ET AL. (2007) e ZHUANG

⁶<http://scantailor.org>

⁷<https://opencv.org>

2. L'acquisizione del testo digitale

E ZHU (2005) hanno integrato con l'output dell'OCR molteplici fonti di informazione letteraria e linguistica, come ad esempio le informazioni riguardanti la struttura fissa della poesia indiana presa in esame o informazioni lessico-semantiche per testi cinesi. Degno di nota è il metodo di PACKER (2011), che usa l'estrazione di informazione da documenti semi-strutturati ai fini della correzione dell'OCR: riconoscere la presenza di una data di nascita, anche se una delle sue cifre è scambiata per un segno di punteggiatura, consente l'autocorrezione dell'errore di OCR.

In ROBERTSON E BOSCHETTI (2017) è illustrato il metodo che si è adottato per acquisire in modo massivo testi greci grazie alla potenza di calcolo messa a disposizione da SHARCNET⁸, la grid canadese per la ricerca accademica. Si rimanda a quel lavoro anche per ulteriori approfondimenti metodologici e bibliografici.

2.2 Metodi per migliorare l'OCR

Si illustrano due diversi metodi per migliorare l'accuratezza dell'OCR. Il primo si basa sulla creazione di profili d'errore per ciascun sistema di OCR grazie all'allineamento dei singoli risultati con rigorose trascrizioni manuali. Questo metodo, che verrà illustrato per esteso, offre la possibilità di un maggiore guadagno nell'accuratezza del risultato finale, ma è dispendioso in termini di tempo da impiegare per eseguire le trascrizioni. Il secondo metodo, che verrà descritto molto più brevemente, non richiede trascrizioni manuali per creare profili d'errore, ma sfrutta la potenza di calcolo di numerosi elaboratori elettronici che lavorano in parallelo.

Per testare il primo metodo sono state usate tre edizioni dei *Deipnosophistae* di Ateneo e un'edizione delle tragedie di Eschilo, estraendo a caso cinque pagine da ciascun esemplare. Tutti i documenti sono stati scaricati da Internet Archive⁹. Gli esemplari di Ateneo appartengono a differenti case editrici e sono fra di loro relativamente lontane nel tempo: MEINEKE (1858) e KAIBEL (1887) sono edizioni

⁸<https://www.sharcnet.ca>

⁹<http://www.archive.org>

Teubner, mentre GULICK (1951) è un'edizione Loeb. Le edizioni delle due case editrici differiscono sensibilmente per tipo di caratteri e quindi sono stati creati diversi modelli (*training sets*) per addestrare i sistemi di OCR al riconoscimento. Le edizioni differiscono inoltre per l'organizzazione dei contenuti: Meineke è privo di apparato critico, Kaibel ha un ricco apparato critico e Gulick ha un apparato critico di dimensioni molto ridotte ma è corredato di note supplementari e di traduzione inglese a fronte.

L'edizione postuma di Eschilo di HERMANN (1852), pubblicata da Weidmann, è priva di apparato critico e ha un tipo di caratteri molto simile alle teubneriane.

Testo greco e apparato critico sono stati trattati separatamente, mentre la traduzione inglese e le note sono state eliminate.

Sistemi di OCR adatti al riconoscimento del Greco antico

Nell'ultimo decennio lo sviluppo delle tecnologie di riconoscimento dei caratteri è stato notevole, soprattutto per il riconoscimento di alfabeti non occidentali e di alfabeti storici. Si illustrano di seguito le caratteristiche di applicazioni di OCR in grado di riconoscere il Greco politonico.

Anagnostis è un'applicazione commerciale non più aggiornata che prende in considerazione spiriti e accenti separatamente dal corpo dei caratteri, facendo aumentare le prestazioni del sistema grazie al numero ridotto di segni da classificare (GATOS ET AL., 2011). Tuttavia Anagnostis non è in grado di riconoscere congiuntamente le sequenze di Greco e di Latino presenti negli apparati critici.

FineReader¹⁰ è un'applicazione commerciale in grado di riconoscere layout complessi e testi poliglotti. È possibile addestrare FineReader con nuovi tipi di caratteri per alfabeti che non siano implementati nativamente, associando immagini di glifi non riconosciuti al codice unicode corrispondente.

¹⁰<http://www.abbyy.com>

2. L'acquisizione del testo digitale

Tesseract¹¹ è un progetto *open source* usato in iniziative di digitalizzazione massiva. Grazie al lavoro di WHITE (2012) attualmente è in grado di riconoscere il Greco politonico in modo molto efficiente¹².

Gamma è un progetto *open source* adattato per il Greco politonico e provvisto di numerosi *training sets* da B. Robertson¹³.

OCROPUS¹⁴ è un progetto *open source* basato su reti neurali dalle prestazioni molto elevate. Può essere addestrato sul Greco politonico e altri alfabeti e, attualmente, è fra i pochi sistemi di OCR in grado di dare buoni risultati anche su documenti a stampa molto antichi¹⁵ e addirittura manoscritti¹⁶.

Gli esempi e gli esperimenti discussi nel seguito sono stati eseguiti con Anagnostis, FineReader e Tesseract.

Addestramento dei singoli sistemi di OCR

Il processo di addestramento al riconoscimento dei caratteri è diviso in due fasi. Nella prima, ciascun sistema di OCR viene addestrato con cinque pagine selezionate casualmente dalle edizioni utilizzate negli esperimenti. Anagnostis e FineReader sono addestrati con gli stessi gruppi di pagine, mentre Tesseract viene addestrato con un gruppo di pagine diverso, al fine di aumentare la possibilità di avere a disposizione campioni di caratteri ignorati dagli altri due sistemi di OCR. In effetti, il problema principale dell'addestramento di FineReader e di Tesseract con il Greco antico è dato dall'elevato numero di caratteri con spiriti e accenti da prendere in considerazione. L'intero set di caratteri pre-combinati per il greco antico contiene più di duecento glifi, ma alcuni sono impiegati con una frequenza molto bassa. Ad esempio, nell'edizione Kaibel di Ateneo, la lettera α (alfa con accento circonflesso, spirito aspro e iota sottoscritto) è presente solo due volte su

¹¹<https://github.com/tesseract-ocr>

¹²I *training sets* sono disponibili sul sito <https://ancientgreekocr.org>.

¹³DALITZ E BASTON (Si veda 2009) e i siti web curati da B. Robertson: <https://github.com/brobertson/rigaudon> e <http://heml.mta.ca/lace>

¹⁴<https://github.com/tmbdev/ocropy>

¹⁵SPRINGMANN E LÜDELING (Si veda 2016).

¹⁶Si veda <https://bit.ly/2uwL44Q>

più di un milione di caratteri. Pertanto, la probabilità che questi caratteri rari siano campionati nei set di addestramento è piuttosto bassa.

Nella seconda fase, gli errori prodotti da ciascun sistema di OCR sono comparati con la cosiddetta *ground truth*, cioè una trascrizione manuale molto accurata per la valutazione dei risultati. In questo modo vengono individuati i pattern d'errore che possono essere corretti tramite la cooperazione di differenti sistemi di OCR. I nuovi *training sets* devono essere identici per tutti i sistemi di OCR. Per l'edizione Weidmann è stato creato un nuovo *training set* di cinque pagine, mentre per le altre edizioni è stato usato il metodo della *k-fold cross validation*, che prevede che vengano usate per il training, a turno, tutte le pagine meno quella usata per il test. La turnazione permette di usare tutte le pagine sia per il training che per il test.

L'output dell'OCR viene allineato alla *ground truth* tramite l'algoritmo di allineamento globale Needleman-Wunsch (FENG E MANMATHAN (si veda 2006) e VAN BEUSEKOM ET AL. (2007)). Gli allineamenti sono ottenuti minimizzando i costi per trasformare una stringa di caratteri nella stringa a essa allineata, tramite l'aggiunta di trattini a indicare i *gap*, quando necessario. In questo modo, gli allineamenti di n-grammi possono essere una coppia di elementi identici (output corretto), una coppia di elementi differenti (errore dovuto a sostituzione), un elemento allineato a un trattino (errore dovuto a inserimento indebito di caratteri) o, infine, un trattino allineato a un elemento trascritto nella *ground truth* (errore dovuto a cancellazione/omissione indebita di caratteri). Dopo l'allineamento, il numero medio di sostituzioni, inserimenti e cancellazioni viene usato per calcolare l'accuratezza media di ciascun sistema di OCR.

I dati relativi agli allineamenti di singoli caratteri, di bigrammi, trigrammi e tetragrammi sono registrati nel file che contiene i *pattern* di errore con una disposizione su quattro colonne: numero di caratteri di cui è costituito l'n-gramma, n-gramma prodotto dall'OCR, n-gramma corrispondente nella *ground truth* e valore di probabilità calcolato dalla formula 2.1.

$$\frac{C(a \rightarrow b)}{C(b)} * \left(\frac{C(b)}{N} \right)^{1/3} \quad (2.1)$$

2. L'acquisizione del testo digitale

Il primo fattore di questo valore esprime la probabilità che, dato un carattere (o un n-gramma) a nell'output dell'OCR, esso rappresenti un carattere (o un n-gramma) b nella *ground truth* (in caso di riconoscimento corretto, a è uguale a b). Tale primo fattore è rappresentato dal numero di occorrenze dell'allineamento corrente, $C(a \rightarrow b)$, diviso per il numero totale di occorrenze del carattere b (o dell'n-gramma) nella *ground truth*, $C(b)$. Il secondo fattore di questo valore è la radice cubica di $C(b)$ divisa per il numero totale di caratteri o n-grammi, N . Questo fattore è uguale per ogni sistema di OCR usato nell'esperimento, perché si basa solo sulla *ground truth* e non sui risultati dell'OCR. Viene fornita la radice cubica di questo valore, in base alla formula 2.6, che verrà illustrata di seguito.

Allineamento multiplo e classificatore bayesiano

Gli algoritmi di allineamento sono comunemente usati in linguistica computazionale per valutare la similarità fra sequenze testuali (KONDRAK, 2002, si veda ad esempio). L'idea di base è l'assegnamento di un costo a ciascuna operazione testuale (sostituzione, inserimento e cancellazione). Per la sostituzione, una matrice di similarità determina costi inferiori per caratteri con forme più simili. Gli inserimenti e le cancellazioni hanno invece costi fissi, mentre l'identità di caratteri non ha costo. Quando due stringhe devono essere allineate, viene creata una matrice con i caratteri della prima stringa disposti in riga e i caratteri della seconda riga disposti in colonna. I costi cumulativi per trasformare la prima stringa nella seconda vengono registrati, muovendosi da sinistra a destra e dall'alto verso il basso lungo le celle della matrice. Seguendo il percorso con il costo minore, a ritroso dall'ultima cella alla prima, è possibile stabilire dove debbano essere inseriti i trattini (cioè i *gap*) per allineare le due stringhe.

Gli output dei tre sistemi di OCR vengono allineati tramite un algoritmo di allineamento multiplo, come illustrato in SPENCER E HOWE (2003). Il principio generale dell'allineamento multiplo progressivo è che si allineano fra di loro prima le coppie di stringhe più simili, vengono fissati i *gap* necessari ad allineare le sequenze e i *gap* supplementari sono aggiunti progressivamente alle sequenze allineate precedentemente, fino al completamento dell'allineamento totale. Al fine di

stabilire quali coppie siano più simili e quindi debbano essere allineate per prime, si dovrebbe costruire un albero filogenetico, ma nel nostro caso di allineamento con sole tre sequenze è sufficiente classificare ciascun sistema di OCR in base all'accuratezza media valutata durante la fase di *training*. Il triplice allineamento è visibile in Fig. 2.1.

L'allineamento in sé non è sufficiente a determinare quale sia il carattere più probabile: anche se due sistemi di OCR concordano, ma sono poco affidabili per l'identificazione di un carattere specifico, il carattere più probabile potrebbe essere fornito dal terzo sistema di OCR che non concorda con gli altri due.

Formalmente, la probabilità che la posizione corrente nella pagina stampata originale e_0 contenga il carattere x , dato che il primo sistema di OCR e_1 fornisce il carattere c_1 , che il secondo sistema e_2 fornisce il carattere c_2 e che il terzo sistema e_3 fornisce il carattere c_3 , è espresso dalla formula:

$$P(e_0 = x | e_1 = c_1, e_2 = c_2, e_3 = c_3) \quad (2.2)$$

dove, in generale, $P(E_0 | E_1, E_2, E_3)$, denota la probabilità a posteriori per l'evento E_0 , data la congiunzione di eventi $E_1 \cap E_2 \cap E_3$.

Ad esempio, l'equazione 2.2 esprime la probabilità che il carattere \tilde{a} si trovi nella posizione corrente sulla pagina a stampa, sapendo che il primo sistema di OCR ha fornito \underline{a} , il secondo il sistema ha fornito \tilde{a} e il terzo sistema ha fornito \acute{a} . Queste probabilità sono dedotte dai dati relativi ai *pattern* di errore registrati durante il processo di addestramento.

Per trovare la probabilità massima tra i tre elementi forniti dai sistemi di OCR, è stato implementato un classificatore bayesiano (più precisamente, un *naive Bayes classifier*). In virtù del teorema di Bayes, l'equazione 2.2 è uguale a:

$$[P(e_1 = c_1, e_2 = c_2, e_3 = c_3 | e_0 = x) * P(e_0 = x)] / P(e_1 = c_1, e_2 = c_2, e_3 = c_3) \quad (2.3)$$

Dato che un classificatore bayesiano si basa sull'assunzione di indipendenza, il primo fattore al numeratore dell'equazione 2.3 può essere riscritto come

$$P(e_1 = c_1 | e_0 = x) * P(e_2 = c_2 | e_0 = x) * P(e_3 = c_3 | e_0 = x) \quad (2.4)$$

Non essendo interessati a trovare il valore della probabilità massima ma soltanto l'argomento x_0 che fornisce la probabilità massima, possiamo omettere il

2. L'acquisizione del testo digitale

denominatore dell'equazione 2.3 e usare la formula seguente:

$$x_0 = \arg \max_x P(e_1 = c_1 | e_0 = x) * P(e_2 = c_2 | e_0 = x) * P(e_3 = c_3 | e_0 = x) * P(e_0 = x) \quad (2.5)$$

Generalizzando, possiamo scrivere l'equazione 2.5 come

$$x_0 = \arg \max_x \prod_{i=1}^n P(e_i = c_i | e_0 = x) * P(e_0 = x)^{1/n} \quad (2.6)$$

dove n è il numero di sistemi OCR usati, e_i è un sistema di OCR specifico e c_i è il carattere fornito in output da quel sistema.

In Fig. 2.1 viene mostrato il risultato della selezione dei caratteri. In diverso colore vengono indicati i caratteri corretti selezionati da Tesseract e Anagnostis, a discapito del carattere riconosciuto da FineReader.

	ἄλλος δ' ἐκείνου παῖς τόδ' ἔργον ἤνυσεν.
FineReader	ἄλλος δ' ἐκείνου_ παῖς τόδ' ἔ_ργον ἠνυσεν.
Tesseract	ἄλλος δ' ἐκείνου* παῖς τόδ' ἔ'ργον ἠνυσεν
Anagnostis	;λλος_ό_ἔχε;του_καῖς τόδ'_ _PYo» ἠνυσιν. -----
Result	ἄλλος δ' ἐκείνου_ παῖς τόδ' ἔ_ργον ἠνυσεν.

Figura 2.1: Allineamento multiplo di tre output dell'OCR

Come spiegato in REYNAERT (2008a) e STEWART ET AL. (2007), la correzione automatica tramite correttori ortografici è da evitare perché spesso è inaffidabile. Il primo suggerimento fornito dal correttore ortografico potrebbe essere errato e parole riconosciute correttamente dal sistema di OCR ma non presenti nel correttore ortografico (in particolare nomi propri) potrebbero essere erroneamente sostituite da parole ortograficamente corrette. Al fine di ridurre questi problemi, si adotta una procedura di correzione ortografica che tenga conto dei risultati di tutti i sistemi di OCR impiegati. In questo modo vengono filtrati soltanto i suggerimenti del correttore ortografico che vengono individuati da una *regular expression* basata sul triplice allineamento.

Spell-checking supportato dall'allineamento multiplo

Al fine di applicare la correzione ortografica automatica si usa la lista di forme flesse generata da Morpheus, l'analizzatore morfologico del Greco antico (CRANE, 1991). La stringa generata dal classificatore bayesiano è analizzata dal correttore ortografico. Quando le parole sono rifiutate dal correttore ortografico perché non sono contenute nella lista delle forme flesse, viene generata una *regular expression* dagli output originali allineati, secondo queste semplici regole: a) i caratteri riconosciuti all'unanimità vengono scritti una volta sola; b) due o tre caratteri riconosciuti in modo diverso vengono scritti fra parentesi quadre; c) i trattini che rappresentano i *gap* sono trasformati in punti di domanda, per indicare nella *regular expression* che il carattere precedente o la coppia di caratteri all'interno delle parentesi quadre sono facoltativi. Ad esempio, dati gli output allineati: a) ἦλασεν, b) ἦλασεν e c) ἦλασεν, la *regular expression* generata è: $/[\grave{\eta}\acute{\eta}]\lambda\alpha\sigma[\epsilon\tau]?v/$. Tutti i suggerimenti forniti dal correttore ortografico sono confrontati con questa *regular expression*, e solo il primo suggerimento che regge il confronto viene selezionato, altrimenti la parola viene lasciata inalterata. Altri esempi sono indicati in Fig. 2.2. Il primo esempio, ἐξερήμωσεν, e l'ultimo, ευφρων, meritano ulteriore attenzione. Nel primo caso, nessun suggerimento fornito dal correttore ortografico regge il confronto con la *regular expression* generata dagli output allineati, così a ragione la parola è lasciata inalterata. Nel secondo caso, ευφρων è una parola malformata in Greco antico perché non ha né spirito né accento. In questo caso, nessun suggerimento del correttore ortografico è supportato dagli output allineati e in questo caso la parola è sì lasciata inalterata, ma erroneamente. Mentre il primo suggerimento del correttore ortografico è scorretto, il terzo è corretto.

2.3 Risultati sperimentali

I risultati vengono valutati confrontando l'accuratezza dei singoli sistemi di OCR con l'accuratezza dell'output allineato, selezionato tramite il classificatore bayesiano e sottoposto al correttore ortografico. Al fine di calcolare l'accuratezza, l'output finale viene allineato con la *ground truth*. Seguendo REYNAERT (2008b),

2. L'acquisizione del testo digitale

FineReader output	RegEx per l'OCR	Spell-checker: suggerimenti	Risultato
ἐξερήμωσεν	ἐξερή[έ?][μι]ωσεν	ἐξερήμωσε, ἐξερήμωσέ, ἐξερήμωσεν	ἐξερήμωσεν
ῶπασεν	[ωοῶ]π[αο]σ[εό]ν	ῶπασεν, ὦπασέν, σπάσεν	ῶπασεν
ἐν'	[εἶ]ν'	ἐν, ἐν' ... ἔν' (34th item)	ἐν'
επάσης	ε?ά?πάσης	πάσης, πάσης ... ἀπάσης (11th item)	ἀπάσης
εὐθυντήριον	[εἶ][ύ]θυντ[ήή]ριον	εὐθυντήριον, εὐθυντήριόν, εὐθυντήρι	εὐθυντήριον
πρώτος	πρ[ώω]τος	πρώτος, πρώτός, πρωτός	πρώτος
Κύρος	[ΚΧΗ][ύι]ρος	Κῦρος, Κῦρός, Κύρος	Κῦρος
ἔθηκε	[εἶ]θηκε	ἔθηκε, ἔθεκέ, θήκε	ἔθηκε
Λυδῶν	[ΛΛ]υδῶν	Λυδῶν, Λιδῶν ... Λυδῶν (6th item)	Λυδῶν
λαὸν	λ[αά][όο]ν	λαὸν, λαόν, Λαίον	λαὸν
ἤλασεν	[ἡῆ]λασ[ετ]?ν	ἤλασεν, ἤλασέν, ἥασεν	ἤλασεν
εὐφρων	ε?ι?[υῶ]φρω?ν	εὐφρων, Εὐφρων, εὔφρων (correct)	εὐφρων

Figura 2.2: Correzione ortografica supportata dall'OCR

l'accuratezza è calcolata come:

$$\frac{\textit{identificazioni}}{\textit{identificazioni} + \textit{sostituzioni} + \textit{inserimenti} + \textit{cancellazioni}} \quad (2.7)$$

cioè come il rapporto tra il numero di caratteri correttamente riconosciuti e la somma di caratteri corretti, sostituiti, inseriti e cancellati nell'OCR.

Il triplice allineamento e il controllo ortografico vincolato dalla *regular expression* forniscono un guadagno di accuratezza, in media, di +2,49% in relazione al miglior sistema di OCR preso singolarmente. Un t-test per ciascun esemplare dimostra che i miglioramenti sono sempre significativi, con $p < 0,05$. L'allineamento senza correzione ortografica fornisce, in media, un miglioramento dell'1%.

Il risultato migliore riguarda la versione Loeb, che è più recente rispetto alle teubneriane prese in considerazione, con un tasso di accuratezza del 99,01%. Se si considera solo il testo senza distinzione tra maiuscole e minuscole, senza segni di punteggiatura e senza spiriti e accenti, l'accuratezza sale a 99,48%. Questo valore è particolarmente importante se si interessati a valutare la *recall* di un sistema

di *emph*text retrieval, dove le parole in Greco antico possono essere cercate senza spiriti e accenti e senza distinzione fra maiuscole e minuscole.

Test sull'apparato critico delle edizioni di Gulick e Kaibel sono stati effettuati senza un addestramento specifico per le note a pie' di pagina, ma con gli stessi set di addestramento applicati al resto della pagina.

L'accuratezza media ottenuta grazie al triplice allineamento è del 92.01%, con un guadagno medio di +3.26% rispetto al miglior sistema di OCR preso singolarmente. Anche sull'apparato critico, il t-test dimostra che i miglioramenti sono significativi, con $p < 0,05$.

Secondo le stime di STEWART ET AL. (2007), l'apparato critico è approssimativamente il 5% della pagina in edizioni con informazioni minime (come le edizioni Loeb) e 14% della pagina, in media, per apparati più informativi (come le edizioni Teubner). Quindi, anche se la quantità di correzioni richieste sull'apparato è superiore, tuttavia la quantità di testo è notevolmente inferiore.

Come anticipato all'inizio della sezione, l'uso del classificatore bayesiano è dispendioso, perché richiede la trascrizione manuale della *ground truth* per creare i profili d'errore. Il secondo metodo non fa uso del classificatore bayesiano e non richiede l'allineamento dell'output di molteplici sistemi di OCR che insistono sulla medesima immagini di testo. È un metodo che si basa sulla semplice stima dei possibili errori commessi da un singolo sistema di OCR e quando tale stima è inferiore a un livello di soglia accettabile, vengono reimpostati i parametri per il *pre-processing* delle immagini (in particolare la soglia di luminosità per eseguire la binarizzazione). Questo metodo è stato usato per il riconoscimento di numerosi volumi scritti in Greco, modificando i parametri di binarizzazione su molteplici nodi della grid canadese e selezionando poi i risultati migliori.

La stima dei possibili errori viene fatta in base al rapporto fra sequenze di caratteri riconosciute come parole greche e sequenze di caratteri totali. Ma per ottenere una stima più approssimata, si può assegnare valore massimo alle sequenze di caratteri riconosciute come parole greche, ma anche valori decrescenti alle sequenze riconosciute come catene sillabiche ben formate (in quanto potrebbero essere nomi propri o varianti morfologiche non registrate nelle liste di forme

2. L'acquisizione del testo digitale

flesse disponibili) e alle sequenze di caratteri che contengono errori soltanto di spiriti o accenti.

2.4 Il processo di correzione manuale

Nel *work-flow* di digitalizzazione i risultati ottenuti tramite procedure automatiche sono mandati a un servizio di *data entry* esterno per le correzioni manuali (si veda CRANE ET AL., 2006). Le ditte di *data entry* lavorano con documenti poliglotti, sui quali non hanno competenze linguistiche specifiche. Per questo motivo, chi lavora al riconoscimento degli errori nelle ditte di *data entry* spesso fa un semplice confronto fra l'immagine della pagina e la forma dei glifi del testo digitalizzato. Si forniscono quindi ai responsabili della correzione i testi formattati con font simili a quelli impiegati nelle edizioni da correggere.

È stato ideato inoltre uno strumento per la correzione manuale dell'OCR, denominato *CoPhiProofReader*¹⁷, che facilita i responsabili alla correzione grazie alla colorazione dei potenziali errori differenziati per tipologia (errore d'accento, catena sillabica ben formata ma non presente nelle liste di forme flesse, etc.). Quando sia disponibile un'altra edizione della stessa opera già digitalizzata, l'OCR viene allineato a questa edizione e vengono prodotte correzioni automatiche, evidenziate per un attento controllo manuale. In Fig. 2.3 le autocorrezioni sono in verde.

2.5 Rimappatura del testo sull'immagine della pagina

La maggior parte dei sistemi di OCR sono in grado di fornire le coordinate dei caratteri riconosciuti sull'immagine originale, ma tale mappatura può richiedere aggiustamenti dopo le correzioni.

Sia il formato PDF che il formato DjVu sono comunemente usati per produrre documenti cercabili con il testo digitalizzato mappato sull'immagine originale. Il

¹⁷Le funzionalità del *CoPhiProofReader* sono illustrate in BOSCHETTI (2014).

2.5 Rimappatura del testo sull'immagine della pagina

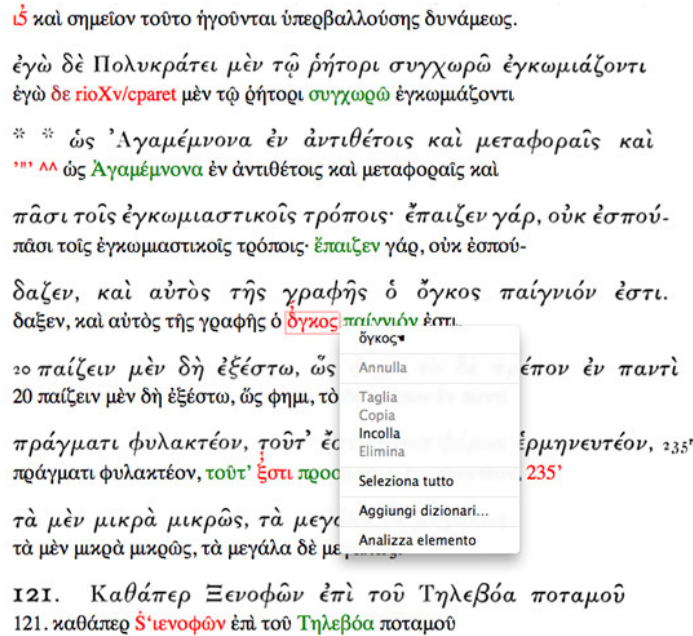


Figura 2.3: CoPhiProofReader

formato PDF è senza dubbio più diffuso; il formato DjVu è nato per gestire in modo ottimale immagini di pagina, grazie a un algoritmo di compressione in grado di separare lo sfondo della pagina (che può essere fortemente compresso) e la scrittura in primo piano (che non può essere compressa eccessivamente, per non perdere leggibilità).

Entrambi i formati supportano l'aggiunta di uno strato (*layer*) di testo sotto l'immagine, in modo che il documento sia cercabile. Ad esempio, GoogleBooks¹⁸ fornisce file .pdf cercabili e Internet Archive¹⁹ fornisce sia file .pdf cercabili che file .djvu.

È vantaggioso mappare il testo sull'immagine della pagina: si preserva il layout originale, anche se il contenuto cercabile è privo di formattazione (*plain text*), e si permette all'utente di verificare se il testo cercabile corrisponde all'immagine da cui deriva.

¹⁸<http://books.google.com>

¹⁹<http://www.archive.org>

2. L'acquisizione del testo digitale

I vari sistemi di OCR hanno capacità diverse rispetto alla mappatura del testo sulle immagini della pagina. Ad esempio, Anagnostis non esporta alcuna informazione sulla mappatura, perché tratta solo file in formato .txt o .rtf. FineReader può esportare file .pdf cercabili, ma difficilmente modificabili. Tesseract, al contrario, esporta i dati in un formato facile da modificare e rimappare sull'immagine.

Tesseract tratta due formati di output: *plain text* e html arricchito con un microformato specifico, chiamato hocr. Un microformato è un compromesso ragionevole tra html puro e sistemi di annotazione più sofisticati, perché estende i tag html con attributi che possono essere elaborati dai parser o ignorati dai browser. (Per un'introduzione generale sui microformati, si veda ALLSOPP, 2007)

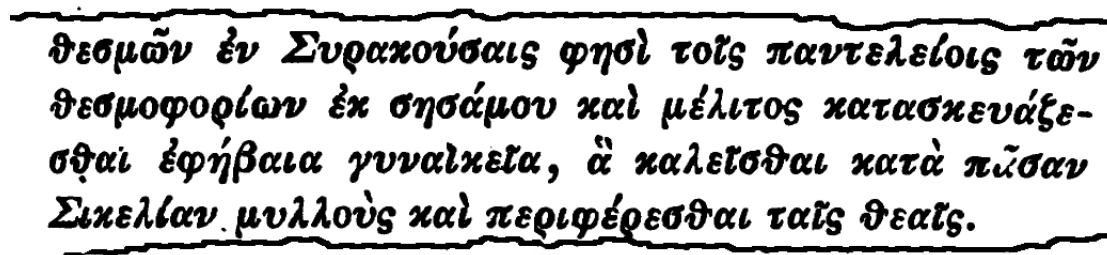


Figura 2.4: Esempio di immagine della pagina

Possiamo considerare la prima riga dell'immagine della pagina in Fig. 2.4. Il testo riconosciuto dal sistema di OCR è *θεσμῶν ἐν Συρακούσαις φησὶ τοῖς παντελείοις το·ν*, con tre errori: un'aggiunta nel caso di *θεσμῶ·ν* invece di *θεσμῶν* e una sostituzione più un'aggiunta nel caso di *το·ν* invece di *τῶν*.

Il microformato hocr memorizza le informazioni di mappatura nell'attributo `title` dei tag html `div`, `p` e `span`. L'attributo `class` descrive il tipo di suddivisione in pagina, paragrafo o linea: `ocr_page`, `ocr_par` o `ocr_line`. In Fig. 2.5 è possibile vedere le coordinate relative alla prima riga di Fig. 2.4. Le coordinate sono fornite carattere per carattere, in una lunga sequenza di numeri relativi all'intera linea.

2.5 Rimappatura del testo sull'immagine della pagina

```
<html>
[...]
<body>
<div class="ocr_page" title="bbox 0 0 1275 1967; image
ath35-194.png">
<p class="ocr_par">

<span class="ocr_line" title="bbox 114 187 1099 234;
bboxes 114 194 140 224, 140 203 154 224, 155 199 175
224, 175 201 198 233, 200 189 228 224, 222 188 227 197,
226 200 249 224, [...]">θεσμῶν ἐν Συρακούσαις φησὶ τοῖς
παντελείοις τον</span><br />
[...]
</p>
[...]
</div>
[...]
</body>
</html>
```

Figura 2.5: Esempio del formato hocr

Il progetto DjVuLibre²⁰ fornisce strumenti per operare sul formato DjVu. In particolare, lo strumento *djvutoxml* estrae il testo ricercabile da un file .djvu. Le coordinate delle parole sulle immagini della pagina, memorizzate nello stesso file .djvu, sono codificate in XML.

Il file .xml contiene metainformazioni sui parametri di scansione, come ad esempio la risoluzione espressa in DPI, e sulla posizione dell'immagine della pagina. La gerarchia delle divisioni del testo nascosto è: colonna della pagina, regione, paragrafo, linea e parola.

Il testo estratto con *djvutoxml* può essere manipolato con un editor di testo e, infine, iniettato nuovamente nel file .djvu con *djvuxmlparser*. Questo strumento legge le metainformazioni relative alle immagini della pagina e rimappa il testo sulla pagina originale.

I file .html arricchiti con il microformato hocr sono facilmente traducibili nel formato DjVuXML. Il risultato è mostrato in Fig. 2.6

²⁰<http://djvulibre.sourceforge.net>

2. L'acquisizione del testo digitale

```
<DjVuXML>
<HEAD>file://localhost/ocr/ath-
meineke/djvu/ath35.djvu</HEAD>
<BODY>
<OBJECT          data=file://localhost/ocr/ath-
meineke/djvu/ath35.djvu
type=image/x.djvu height=1967 width=1275 usemap=-
ath35-194.djvu>
<PARAM name=DPI value=100/>
<PARAM name=GAMMA value=2.200000/>
<PARAM name=PAGE value=ath35-194.djvu/>
<HIDDENTEXT><PAGECOLUMN><REGION><PARAGRAPH>
<LINE>
<WORD coords=114,194,249,233>θεσμῶν</WORD>
<WORD coords=271,187,312,233>ἐν</WORD>
<WORD coords=332,187,574,233>Συρακούσais</WORD>
<WORD coords=593,187,678,234>φησὶ</WORD>
<WORD coords=697,187,767,234>τοῖς</WORD>
<WORD coords=787,187,1007,234>Παντελείους</WORD>
<WORD coords=1027,187,1078,234>τον</WORD>
</LINE>
[...]
</PARAGRAPH></REGION></PAGECOLUMN></HIDDENTEXT>
</OBJECT>
</BODY>
</DjVuXML>
```

Figura 2.6: Esempio di formato DjVuXML

Se il risultato dell'OCR è stato corretto con il *CoPhiProofReader*, che accetta il formato hocr, le coordinate delle parole vengono conservate. Solo in caso di divisione di una parola in due o di unione di due parole in una sono necessari aggiustamenti alle coordinate. Se il risultato dell'OCR è stato invece corretto con altri strumenti, le coordinate potrebbero essere completamente perdute oppure potrebbero essere conservate le sole coordinate di riga, ma non di parola. In questo caso sono necessarie procedure di riallineamento fra la versione originaria con le coordinate di parola e la nuova versione corretta.

In Fig. 2.7 è possibile vedere l'esempio di un file DjVu cercabile, dopo che il testo corretto è stato rimappato sull'immagine.

2.5 Rimappatura del testo sull'immagine della pagina

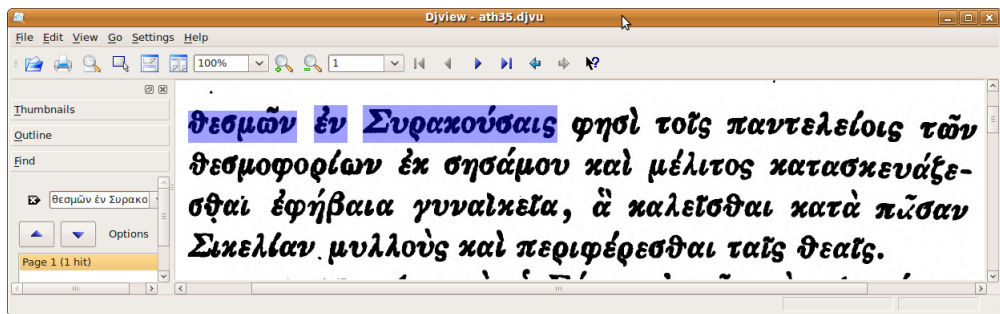


Figura 2.7: Text retrieval in Djview

3

Analisi di apparati critici e repertori di congetture

Obiettivi e limiti

Si cercherà in questo capitolo di affrontare gli aspetti teorici e di mettere in luce i problemi relativi alla marcatura di apparati critici¹ e repertori di congetture al fine di ampliare il database basato sull'edizione di riferimento con i dati relativi a varianti e congetture. Basandosi soltanto sulle informazioni contenute nei repertori, che spesso sono lacunose, soprattutto per quanto riguarda colometria e punteggiatura, il risultato ottenuto con la marcatura può considerarsi come uno strumento di ausilio per successive indagini sui testi presi in esame dagli autori dei repertori stessi e non come uno strumento che riproduca nella sua totalità le informazioni contenute nei testi di origine.

Il testo stabilito dall'editore è il risultato di un processo di selezione di varianti e congetture basato sulla conoscenza della trasmissione di quel testo, dello stile dell'autore e del contesto storico-linguistico. Ogni nuova edizione stabilisce un testo diverso dalle precedenti: l'editore può attenersi alla paradosi, può accogliere congetture brillanti di altri studiosi o può formularne di nuove. Le sue scelte sono

¹Aspetti teorici di rilievo sono stati affrontati da FROGER (1968), da MORDENTI (2001) e da BUZZETTI (1999), che ragiona sull'adeguatezza dei linguaggi di *markup* per rappresentare testi con varianti. Si vedano anche BOZZI ET AL. (1986).

3. Analisi di apparati e repertori

in parte individuali e in parte influenzate dalla tradizione di studi e dal paradigma ermeneutico della scuola a cui appartiene.

Chi usa un *corpus* letterario antico deve avere la consapevolezza che i testi degli autori sono filtrati dai giudizi (e talvolta dai pregiudizi) degli editori: non si può studiare un fenomeno linguistico o stilistico se quel fenomeno è deformato dalle scelte dell'editore stesso. Un caso esemplare è lo studio delle ripetizioni: il paradigma tardo-ottocentesco tendeva a considerare molte ripetizioni a breve termine come errori commessi dai copisti, perciò gli editori preferivano omettere o sostituire tali ripetizioni tramite ricostruzioni congetturali, spesso del tutto arbitrarie. Il nuovo paradigma, invece, riconosce come genuino questo dispositivo stilistico: il risultato inaspettato di PICKERING (2000) è che i copisti erano allenati a togliere ripetizioni, anziché ad aggiungerle in modo indebito. Tuttavia, se si vuole sostenere questa posizione tramite analisi stilistiche condotte su *corpora* digitali, ci si trova a dover ignorare numerose ripetizioni presenti nei manoscritti, proprio perché sono state soppresse dagli editori moderni nelle loro edizioni. Di conseguenza, le concordanze basate su tali edizioni non consentono lo studio del fenomeno nella sua reale dimensione. In assenza di una nuova collazione, sarebbero le informazioni presenti negli apparati critici a dare contezza del fenomeno.

Come detto più volte nei capitoli precedenti, le raccolte più complete di testi greci e latini sono basate su autorevoli edizioni moderne, ma mancano di apparato critico e quindi non contengono informazioni su varianti testuali attestate nei manoscritti o congetture formulate dagli studiosi. Grazie a iniziative delle singole biblioteche, come (fra le altre) la Biblioteca Nazionale Marciana², la Biblioteca Medicea Laurenziana³ o la Biblioteca Apostolica Vaticana⁴ per i manoscritti e a iniziative come Internet Archive⁵ per i testi a stampa, attualmente è possibile accedere a una grande quantità di documenti digitali per la consultazione delle fonti primarie e della letteratura secondaria. Questo agevola notevolmente lo

²<https://marciana.venezia.sbn.it>

³<https://www.bmlonline.it>

⁴<https://www.vaticanlibrary.va>

⁵<http://www.archive.org>

studioso, ma si deve comunque distinguere fra testo digitalizzato, *machine actionable* e quindi cercabile, copiabile, trattabile come unità testuale su cui effettuare statistiche, e mera immagine digitale del testo, leggibile da un essere umano ma non trattabile da un calcolatore. La maggior parte delle immagini di manoscritti e di edizioni a stampa disponibili negli archivi digitali non è ancora associata ai relativi testi digitalizzati, sufficientemente affidabili per la ricerca testuale con sistemi di *text retrieval*. Resta quindi ancora parzialmente valida l'osservazione di DEGANI (1992), secondo cui il filologo che usa strumenti computazionali deve poi verificare i suoi risultati su edizioni a stampa (sia in biblioteca, come negli anni Novanta del secolo scorso, sia in una biblioteca digitale che mette a disposizione le immagini delle pagine). Il testo di un'edizione critica senza apparato, infatti, è privo di valore scientifico, non essendo possibile distinguere fra varianti attestate nei manoscritti e ricostruzioni congetturali.

Eppure anche l'apparato critico è una selezione. Se il testo accolto dall'editore è soggettivo nelle sue sostituzioni, l'apparato critico è soggettivo nelle sue omissioni. L'apparato critico registra varianti con i *sigla* dei manoscritti e congetture con riferimenti bibliografici, ma può essere considerato un'antologia e non un repertorio esaustivo di informazioni. Solo repertori di collazioni e repertori di congetture possono rivendicare la completezza, anche se il primo è limitato dal numero di manoscritti esaminati e il secondo dal numero di edizioni a stampa, commenti e articoli consultati.

Per i motivi spiegati, l'interesse ad arricchire i *corpora* letterari con varianti e congetture sta crescendo e catalizza l'attenzione di diversi gruppi di ricerca; tra molti altri, l'*Homer Multitext Project*⁶ e *Musisque Deoque*⁷. Per un approfondimento teorico sulla relazione tra testi e apparati nelle edizioni digitali, si veda FROGER (1968), BOZZI ET AL. (1986), BUZZETTI (1999), MORDENTI (2001) e BOZZI (2004a).

I *corpora* digitali di lingue antiche possono essere estesi non solo con varianti e congetture, ma anche con annotazioni linguistiche, metriche e stilistiche (lemma-

⁶http://www.chs.harvard.edu/publications.sec/homer_multitext.ssp

⁷<http://www.mqdq.it>

3. Analisi di apparati e repertori

tizzazione, annotazione morfo-sintattica, analisi metrica, etc.). Lo stesso *corpus* annotato può essere esteso in modo asincrono da gruppi e istituzioni indipendenti, dando luogo a problemi di manutenzione, compatibilità, riferimenti incrociati, etc. Per questo è bene che i *corpora* siano gestiti all'interno di infrastrutture di ricerca per le risorse linguistiche, come CLARIN⁸.

È difficile determinare l'unità di base delle varianti e delle congetture. Se l'attenzione è focalizzata su come la *varia lectio* ha origine da un punto di vista paleografico, il singolo carattere è l'unità di base più adatta. Ma dal punto di vista linguistico e stilistico, l'unità di base è invece la parola, che può essere raggruppata in unità superiori (ad esempio il verso che contiene la variante), o divisa in unità inferiori (ad esempio le sue sillabe).

I repertori sottoposti a marcatura

L'edizione critica elettronica dovrebbe porsi come obiettivo il superamento della netta separazione fra testo accolto e apparato. Se l'organizzazione delle informazioni sulla pagina cartacea è vincolato allo spazio bidimensionale dove collocare il testo, l'edizione elettronica può disporre di uno spazio multidimensionale, che permette di visualizzare, o video o sulla carta, molteplici sezioni dell'iperspazio dove risiede concettualmente il testo digitale. L'edizione di riferimento non è più intesa, quindi, come il testo accolto su cui costruire indici, concordanze e altri strumenti ausiliari; diventa invece il mezzo che fornisce un sistema coerente di posizioni testuali, indipendentemente dal contenuto dell'unità testuale che andrà a occupare quella posizione. L'edizione di riferimento offre quindi l'ordine testuale convenzionale, la successione di caselle vuote numerate progressivamente, rispetto alle quali collocare ciascuna unità testuale proveniente da diversi repertori. Nel nostro caso essi sono costituiti dagli apparati critici di MURRAY (1955) e di WEST (1998), da cui si estraggono anche i dati relativi alle varianti testuali, e poi i repertori di congetture propriamente detti di WECKLEIN (1893), DAWE (1965) e l'appendice contenuta negli *Studies* di WEST (1990).

⁸<https://www.clarin.eu>

La ricostruzione del contesto

Ciascun repertorio usa una differente edizione critica⁹ come proprio sistema di riferimento, cioè per la numerazione dei versi e per determinare il contesto nel quale collocare la congettura. Per la ricostruzione di tale contesto, costituito dalle due o tre parole precedenti e seguenti la congettura stessa, è necessario procedere alla collazione completa delle rispettive edizioni di riferimento, in modo tale da usare la seconda edizione di Murray per la matrice generale delle posizioni testuali, ma anche le edizioni di Wecklein e di West per la ricostruzione dei rispettivi contesti.

La struttura dei repertori

La filologia computazionale, sfruttando tecniche già ampiamente utilizzate dai programmatori per individuare le differenti versioni dei listati di uno stesso programma, ha fornito validi strumenti di ausilio al processo di collazione automatica. E' possibile infatti, ormai con una certa facilità, individuare meccanicamente le differenze fra due o più testi digitalizzati, in modo da segnalare dove l'uno presenti aggiunte, sottrazioni, sostituzioni o trasposizioni di testo rispetto all'altro. Tuttavia, questo tipo di procedimento non è applicabile nel nostro caso. Infatti le informazioni che ci interessano, cioè le lezioni congetturali raccolte dai curatori dei repertori, non si trovano soltanto nelle differenti edizioni del testo preso in esame, ma anche in commentari, in saggi e in articoli di riviste di filologia.

I diversi curatori dei repertori hanno organizzato le informazioni relative alle congetture in modi differenti: dalle indicazioni a volte vaghe di Wecklein, alle notizie bibliografiche estese di Dawe, si è giunti al sistema sintetico e rigoroso di indicazioni bibliografiche di West.

La creazione di indici e concordanze a partire da tali repertori richiede quindi un'operazione di riorganizzazione, di normalizzazione e di integrazione delle informazioni. Per quanto ciascun curatore si sia attenuto a criteri coerenti di

⁹Il repertorio di Wecklein costituisce un'appendice alla sua stessa edizione: WECKLEIN (1885); Dawe si basa sull'edizione di Murray e West fa riferimento alla propria edizione.

3. Analisi di apparati e repertori

registrazione delle congetture, il fatto di rivolgersi a un pubblico specialistico ha permesso il ricorso a forme di registrazione facilmente interpretabili dal lettore ma difficilmente analizzabili dal calcolatore.

Prendendo un caso semplice, al v. 815 (818 Wecklein)

κρηπίς ὑπεστιν, ἀλλ' ἔτ' ἐκπιδύεται.

il repertorio di Wecklein riporta: *πάρεστιν* Rauchenstein. *ἐκμαιεύεται* Herm., *ἐξαδρύνεται* Oberdick, *ἐκπληθύεται* Paley. Come è evidente, *πάρεστιν* è congettura per *ὑπεστιν*, mentre *ἐκμαιεύεται*, *ἐξαδρύνεται* ed *ἐκπληθύεται* sono congetture per *ἐκπιδύεται*; tuttavia questo è chiaro al lettore in base a considerazioni morfosintattiche non immediatamente disponibili al calcolatore. In altri casi il curatore del repertorio riporta una o più parole del contesto in cui va inserita la congettura e in altri casi ancora, dove l'intervento testuale sia più complesso, viene riportato addirittura un contesto ampio quanto un intero periodo, a volte costituito da più versi. In questo modo si rende impossibile l'applicazione di procedure automatiche semplici, in grado di individuare la posizione testuale precisa di ciascuna parola formulata all'interno di una congettura. La marcatura viene condotta quindi con un procedimento ampiamente manuale; alcune operazioni di *markup* sono tuttavia effettuabili dal calcolatore sulla base di accorgimenti adottati già in fase di digitalizzazione dei testi, come l'introduzione di a capo per ciascuna notizia dei repertori e l'uso della codifica *unicode*, che consente, all'interno di *files* di solo testo, l'immediata individuazione dei set di caratteri latini rispetto ai set di caratteri greci e quindi l'isolamento automatico delle parole greche contenute nei repertori.

3.1 Le linee guida della Text Encoding Initiative

La *Text Encoding Initiative*¹⁰ propone tre differenti metodi per la marcatura delle notizie d'apparato: il *location-referenced method*, il *double end-point attachment*

¹⁰Il sito ufficiale si trova all'indirizzo <<http://www.tei-c.org>>, dove è possibile trovare tutti i materiali utili e la documentazione, in particolare le *Guidelines*, che sono conformi al volume di SPERBERG-MCQUEEN E BURNARD (2002) e contengono aggiornamenti.

3.1 Le linee guida della Text Encoding Initiative

method e il *parallel segmentation method*. Il primo è molto adatto per la digitalizzazione completa di una singola edizione critica cartacea. Nel testo viene marcata la sequenza interessata e nell'apparato si fa riferimento a tale blocco.

Ad esempio, se si volessero marcare i vv. 813-817 dell'edizione di Wecklein¹¹:

βωμοὶ δ' αἴστοι, δαιμόνων θ' ἰδρύματα
πρόρριζα φύρδην ἐξανέστραπται βάρων.
τοιγὰρ κακῶς δράσαντες οὐκ ἐλάσσονα
πάσχουσι, τὰ δὲ μέλλουσι, κοῦδέπω κακῶν

813 δαιμόνων θ' recc.: δαιμόνων M.

817 ἐκπιδύεται Schuetz: ἐκπαιδεύεται M.

il risultato sarebbe:

```
1 <text>
2 <body>
3 <!-- ... -->
4 <div n="persae" type="tragedy">
5 <l n="300">βωμοὶ δ' αἴστοι, δαιμόνων θ' ἰδρύματα</l>
6 <l>πρόρριζα φύρδην ἐξανέστραπται βάρων.</l>
7 <l>τοιγὰρ κακῶς δράσαντες οὐκ ἐλάσσονα</l>
8 <l>πάσχουσι, τὰ δὲ μέλλουσι, κοῦδέπω κακῶν</l>
9 <l n="301">κρηπὶς ὑπεστυν, ἀλλ' ἔτ' ἐκπιδύεται.</l>
10
11 <!-- ... -->
12
13 <app loc="300">
14 <rdg wit="recc.">δαιμόνων θ'</rdg>
15 <rdg wit="M">δαιμόνων</rdg>
16 </app>
17 <app loc="persae 301">
18 <rdg resp="Schuetz">ἐκπιδύεται</rdg>
19 <rdg wit="M">ἐκπαιδεύεται</rdg>
20 </app>
```

Se il primo metodo fa riferimento alla linea di testo, senza indicazioni più precise, il *double end-point attachment method* permette un riferimento preciso fra la porzione di testo e la notizia di apparato. È possibile in questo modo non solo riprodurre l'edizione cartacea nella sua veste grafica, ma anche, se desiderato, sostituire dinamicamente il contenuto dell'apparato inserendolo nella linea di

¹¹Si segnala che, per stessa ammissione di Wecklein, fra l'edizione e l'apparato nella numerazione vi è la discrepanza di un verso, quindi il v. 817 dell'edizione corrisponde al v. 818 dell'apparato.

3. Analisi di apparati e repertori

testo, ad esempio per lavorare a una nuova edizione critica. In questo secondo caso l'esempio visto sopra assumerebbe la forma:

```
1 <text>
2 <body>
3 <!-- ... -->
4 <div n="persae" type="tragedy">
5 <l n="300">βωμοὶ δ' ἄιστοι, <anchor id="300.0">δαιμόνων
6 <anchor id="301.1">θ' ἰδρύματα
7 </l>
8 <l>πρόρριζα φύρδην ἐξανέστραπται βάθρων.</l>
9 <l>τοιγὰρ κακῶς δράσαντες οὐκ ἐλάσσονα</l>
10 <l>πάσχουσι, τὰ δὲ μέλλουσι, κοῦδέπω κακῶν</l>
11 <l n="301">κρηπὶς ὕπεστιν, ἀλλ' ἔτ'
12 <anchor id="301.0">ἐκπιδύεται.<anchor id="301.1">
13 </l>
14
15 <!-- ... -->
16
17 <app from="300.0" to="300.1">
18 <rdg wit="recc.">δαιμόνων θ'</rdg>
19 <rdg wit="M">δαιμόνων</rdg>
20 </app>
21 <app from="301.0" to="301.1">
22 <rdg resp="Schuetz">ἐκπιδύεται</rdg>
23 <rdg wit="M">ἐκπαιδεύεται</rdg>
24 </app>
```

Nell'ultimo metodo le notizie d'apparato sono considerate come alternative l'una all'altra, non rispetto al testo di riferimento. Testo e apparato confluiscono in una unica struttura:

```
1 <text>
2 <body>
3 <!-- ... -->
4 <div n="persae" type="tragedy">
5 <l>βωμοὶ δ' ἄιστοι,
6 <app>
7 <rdg wit="M">δαιμόνων</rdg>
8 <rdg wit="recc.">δαιμόνων</rdg>
9 </app>
10 θ' ἰδρύματα
11 <l>πρόρριζα φύρδην ἐξανέστραπται βάθρων.</l>
12 <l>τοιγὰρ κακῶς δράσαντες οὐκ ἐλάσσονα</l>
13 <l>πάσχουσι, τὰ δὲ μέλλουσι, κοῦδέπω κακῶν</l>
14 <l n="301">κρηπὶς ὕπεστιν, ἀλλ' ἔτ'
15 <app>
16 <rdg wit="M">ἐκπαιδεύεται</rdg>
17 <rdg resp="Schuetz">ἐκπιδύεται</rdg>
18 </app>
19 </l>
20
21 <!-- ... -->
22
23 <app loc="persae 300">
24 <rdg wit="recc.">δαιμόνων θ'</rdg>
25 <rdg wit="M">δαιμόνων</rdg>
```

3.1 Le linee guida della Text Encoding Initiative

```
26 </app>
27 <app loc="persae 301">
28 <rdg resp="Schuetz">ἐκπιδύεται</rdg>
29 <rdg wit="M">ἐκπαιδύεται</rdg>
30 </app>
```

Questo metodo tuttavia non permette, nel caso di varianti o congetture formate da più parole, che le strutture si sovrappongano ma soltanto che siano contenute l'una nell'altra.

Purtroppo tutti questi metodi sono ottimali per una quantità limitata e già predeterminata di notizie d'apparato ma presentano alcune carenze per un lavoro come il nostro, inteso ad estendersi con un numero elevato di notizie che insistono tutte sulle medesime posizioni testuali, che hanno numerose possibilità di sovrapposizione, etc. Il *double end-point attachment method* sembra il più adatto ai nostri scopi, in quanto combina la precisione nello stabilire il punto di inserimento delle alternative alla possibilità di parziali sovrapposizioni fra le sequenze testuali. In un sistema dove non sia predeterminato il numero delle notizie d'apparato da collegare al testo, l'accorgimento da prendere è quello di marcare con `<anchor xml:id="..." />` tutte le posizioni testuali con numeri progressivi. Tuttavia, dato che i testi vengono marcati per poter essere successivamente analizzati linguisticamente, e dato che l'unità linguistica più naturale ai fini della lemmatizzazione, dell'analisi morfo-sintattica e dell'analisi semantica è la parola, si preferisce usare il marcatore TEI `<w>...</w>` anziché introdurre ancora prima e dopo ogni singola parola.

Data la dispendiosità della marcatura manuale, si è proceduto a marcare manualmente campioni di testo con il *location-referenced method*, inserendo le informazioni necessarie all'applicazione del *double end-point attachment method* in automatico, avendo la possibilità di usare come ancore gli identificatori univoci delle parole, generati in automatico sfruttando i valori delle posizioni nel testo. Come sarà spiegato nel capitolo successivo, la marcatura manuale è servita come studio per elaborare procedure di marcatura automatica.

3. Analisi di apparati e repertori

3.2 Marcatura dei repertori di congetture

Le principali operazioni da compiere per marcare le informazioni pertinenti al nostro lavoro sono le seguenti:

1. suddivisione dei campioni di testo dei repertori in sezioni e sottosezioni costituite dai seguenti elementi TEI:

app: porzione di testo compreso fra il riferimento al verso al quale appartengono le congetture, indicato su carta in neretto, e il riferimento successivo;

rdg: porzione di testo contenuta nell'*app* e costituita da una o più parole oppure operazioni testuali (come cancellazione o omissione intenzionale), alternative al testo dell'edizione di riferimento del repertorio oppure ad altre congetture presenti nello stesso *app* e affiancate da una indicazione bibliografica più o meno accurata – l'attributo *n* indica la posizione della prima parola della congettura rispetto all'edizione di riferimento principale;

w: porzione di testo della congettura rappresentata da una forma in lingua greca oppure da una operazione testuale (come cancellazione o omissione intenzionale, segnalazione di una lacuna, etc.) – l'attributo *n* indica, quando necessario, la posizione precisa rispetto all'edizione di riferimento principale;

span: uno *span* può contenere informazioni di vario genere, come il numero di verso, il nome del responsabile della congettura ed eventuali indicazioni bibliografiche relative al luogo citato, oppure l'indicazione di *dramatis personae*, punteggiatura, colometria, quando presenti nel repertorio. Può inoltre marcare informazioni discorsive per indicare operazioni testuali, come *delet*, *omittit*, *iterat* oppure giudizi, come *suspectat*, *dubitanter correxi*, etc. Grazie all'attributo *type* è possibile usare l'informazione contenuta nello *span* per trasferirla negli attributi di altri marcatori. Ad esempio, il nome del responsabile è inserito in *span* di tipo *rdg_resp*, in modo da riportare automaticamente (tramite un programma, uno script o un foglio di stile) l'informazione nell'attributo *resp* di *rdg*, come indicato dalle linee guida della TEI.

Si può vedere un esempio per capire meglio. Si prenda in considerazione il v. 428 (431 secondo la numerazione di Wecklein):

3.2 Marcatura dei repertori di congetture

ἕως κελαινῆς νυκτὸς ὄμμ' ἀφείλετο

e si vedano le relative informazioni ricavate dal repertorio di Wecklein:

431 κελαινῆ νύξ τόδ' ὄμμ' olim CGHaupt. ὄρφν' ἀφείλετο Volckmar, οἶμ' ἀφεί-
λένιν Heimsoeth, ἄρμ' ἀφείλετο Oberdick (mallem ἄρμ' ἠπείγετο). ἀφείλετο
Pauw, ἀφίκετο Blomf., ἐφαίνετο Wiel.

La marcatura corrispondente assume questo aspetto:

```
1 <listApp xml:base="urn:cts:greekLit:tlg0085.tlg002.opp-grc6:">
2 <app loc="428">
3 <span type="app_loc">431.</span>
4 <rdgGrp>
5 <rdg n="2165" resp="olim Haupt">
6 <w>κελαινῆ</w>
7 <w>νύξ</w>
8 <w n="2166.1">τόδ'</w>
9 <w n="2167">ὄμμ'</w>
10 <span type="rdg_resp">olim CGHaupt</span>
11 <span type="rdg_grp_sep">.</span>
12 </rdg>
13 </rdgGrp>
14 <rdgGrp>
15 <rdg n="2167" resp="Volckmar">
16 <w>ὄρφν'</w>
17 <w>ἀφείλετο</w>
18 <span type="rdg_resp">Volckmar</span>
19 <span type="rdg_sep">,</span>
20 </rdg>
21 <rdg n="2167" resp="Heimsoeth">
22 <w>οἶμ'</w>
23 <w>ἀφείλέ</w>
24 <w n="2168.1">νιν</w>
25 <span type="rdg_resp">Heimsoeth</span>
26 <span type="rdg_sep">,</span>
27 </rdg>
28 <rdg n="2167" resp="Oberdick">
29 <w>ἄρμ'</w>
30 <w>ἀφείλετο</w>
31 <span type="rdg_resp">Oberdick</span>
32 </rdg>
33 <rdg n="2167" resp="Wecklein">
34 <span type="lpar">(</span>
35 <span type="judgment">mallem</span>
36 <w>ἄρμ'</w>
37 <w>ἠπείγετο</w>
38 <span type="rpar">)</span>
39 <span type="rdg_grp_sep">.</span>
40 </rdg>
41 </rdGrp>
42 <rdgGrp>
43 <rdg n="2168" resp="Pauw">
44 <w>ἀφείλετο</w>
45 <span type="rdg_resp">Pauw</span>
46 <span type="rdg_sep">,</span>
```

3. Analisi di apparati e repertori

```
47     </rdg>
48     <rdg n="2168" resp="Blomfield">
49       <w>ἀφίκετο</w>
50       <span type="rdg_resp_abbrev">Blomf.</span>
51       <span type="rdgSep">,</span>
52     </rdg>
53     <rdg n="2168" resp="Wiel">
54       <w>ἐφαίvero</w>
55       <span type="rdg_resp">Wiel</span>
56       <span type="rd_grp_sep">.</span>
57     </rdg>
58   </rdgGrp>
59 </app>
60 </listApp>
```

Tutte le *app* sono contenute in *listApp*, che possiede l'attributo *xml:base* ad indicare che l'edizione di riferimento principale è MURRAY (1955), grazie all'uso del CTS-URN fornito dal *Perseus Catalog*¹². L'attributo *loc* di *app* indica il primo verso a cui si fa riferimento nell'edizione di Murray, e serve soltanto a facilitare la lettura, con un *editor* di testo, del documento marcato. Il sistema di numerazione dei versi per la creazione degli indici e delle concordanze, infatti, si affida a una tabella di corrispondenza fra numero di verso e posizione della prima parola del verso. In questo modo è possibile passare velocemente da un sistema di numerazione a un altro, ad esempio di una edizione più recente, aggiornando solamente la tabella di corrispondenza.

L'attributo *n* di ciascuna congettura (*rdg*) indica la posizione della prima parola rispetto all'edizione di Murray; le parole (*w*) possiedono il proprio attributo *n* solo se la numerazione progressiva, a partire dalla prima parola della congettura, è diversa dalla numerazione dell'edizione di riferimento. Ciò si verifica principalmente a causa di inserzione di testo fra due posizioni che sono contigue nell'edizione di riferimento. È il caso, nell'esempio, di <w n="2168.1">νν</w>, parola che deve essere inserita fra ἀφείλέ, che nella congettura di Heimsoeth sostituisce l'ultima parola del verso (posizione 2168), e la prima parola del verso successivo (posizione 2169)¹³.

¹²<http://catalog.perseus.org>

¹³Questo esempio è semplificato, in quanto nella pratica si devono dare anche informazioni ulteriori per indicare se la parola viene aggiunta alla fine del verso corrente o all'inizio del verso successivo. In caso di cancellazione di testo, vengono usati elementi <w> vuoti. In caso di trasposizione gli elementi vuoti, dotati di un proprio attributo *xml:id*, sono usati in congiunzione

3.2 Marcatura dei repertori di congetture

Gli *span*, come si vede, sono quasi sempre dotate dell'attributo *type*, al fine di riusare l'informazione, come nel caso degli *span* di tipo *rdg_resp*, il cui contenuto viene appunto riusato nell'attributo *resp* della *rdg*.

Un caso che richiede particolare attenzione è rappresentato dalle strutture nidificate, cioè da congetture formulate nel contesto di altre congetture e indicate dai curatori nel modo più sintetico possibile. Un esempio, tratto dal repertorio di Dawe:

239 διὰ χερὸς (seu χερῶν) σφισὶν πρέπει B (excepto χερῶν, iam Wellauer)
διὰ χερὸς λαοῖς Weil (iam Enger) κέρως pro χερῶν Broadhead

e la relativa marcatura:

```
1 <app loc="239">
2 <span type="app_loc">239</span>
3 <rdgGrp>
4 <rdg n="1138" resp="Blaydes Wellauer">
5 <w>διὰ</w>
6 <w>χερὸς</w>
7 <app>
8 <rdg>
9 <span type="lpar">(</span>
10 <span>seu</span>
11 <w n="1139">χερῶν</w>
12 <span>)</span>
13 </rdg>
14 </app>
15 <w n="1140">σφισὶν</w>
16 <w>πρέπει</w>
17 <span type="rdg_resp_abbrev">B</span>
18 <span type="lpar">(</span>
19 <span>excepto χερῶν</span>
20 <span>,</span>
21 <span>iam</span>
22 <span>Wellauer</span>
23 <span type="rpar">)</span>
24 </rdg>
25 <rdg n="1138" resp="Weil Enger">
26 <w>διὰ</w>
27 <w>χερὸς</w>
28 <w>λαοῖς</w>
29 <span>Weil</span>
30 <span type="lpar">(</span>
31 <span>iam</span>
32 <span type="rdg_resp">Enger</span>
33 <span type="rpar">)</span>
```

con elementi *<w>* dotati dell'attributo *corresp*; le trasposizioni infatti sono sempre trattate a livello di *<w>* e non a livello di *rdg*. È importante ricordare che le operazioni testuali registrate (sostituzione, addizione, cancellazione e trasposizione) hanno un significato puramente funzionale alla creazione di indici e concordanze, e non hanno alcun valore interpretativo.

3. Analisi di apparati e repertori

```
34     </rdg>
35 </rdgGrp>
36 <rdgGrp>
37   <rdg n="1139">
38     <w>κέρως</w>
39     <span>pro χερῶν</span>
40     <span type="rdg_resp">Broadhead</span>
41   </rdg>
42 </rdgGrp>
43 </app>
```

Come si vede, la congettura *διὰ χερῶν σφισὶν πρέπει* è indicata quale alternativa dello stesso B(laydes) alla congettura *διὰ χερὸς σφισὶν πρέπει*.

3.3 Elaborazione dei documenti marcati

I documenti marcati manualmente vengono sottoposti a un programma che ha lo scopo di estrarre i dati e inserirli all'interno di un database. Molti linguaggi di programmazione offrono la possibilità di elaborare i dati di un documento XML, anche senza far ricorso ai fogli di stile XSL(T).

Il *World Wide Web Consortium*¹⁴ propone come standard per l'elaborazione dei documenti un modello, DOM (*Document Object Model*), che si basa sulla mappatura della struttura ad albero di un documento XML sulla struttura ad albero della gerarchia di classi utilizzata comunemente nei linguaggi di programmazione a oggetti. DOM è il modello sul quale si basa la maggior parte dei motori XSL(T), che effettuano l'elaborazione dei dati contenuti in un documento XML seguendo le direttive di fogli di stile. I maggiori linguaggi di programmazione affiancano a DOM modelli alternativi per processare i documenti XML, fra cui SAX (*Simple API for XML*). SAX è un modello non gerarchico ma *event driven*. Gli eventi che guidano il flusso del programma sono legati al tipo di dato XML. Per esempio, l'apertura e la chiusura di un *tag*, incontrati durante la lettura sequenziale del file XML da parte del programma, sono considerati nel modello SAX come *events*, al verificarsi dei quali il programma deve rispondere invocando le relative procedure. Questo permette di eludere in modo molto agile la struttura ad albero apparen-

¹⁴Tutte le specifiche del W3C sono consultabili sul sito ufficiale all'indirizzo: <http://www.w3c.org>.

3.3 Elaborazione dei documenti marcati

temente imposta da SGML/XML, facendo ampio ricorso ai *milestone elements* proposti dalle direttive TEI¹⁵, e dando quindi la possibilità di gestire facilmente i casi di *overlapping*¹⁶. Per esempio, dato un frammento XML di questo tipo:

```
1 <my-tag id="1" pos="start"/>
2   a b c
3 <my-tag id="2" pos="start"/>
4   d e f
5 <my-tag id="1" pos="end"/>
6   g h i
7 <my-tag id="2" pos="end"/>
```

si può creare una procedura che inserisca in variabili differenti il testo incontrato fra gli eventi di apertura/chiusura di un marcatore vuoto con un determinato *id* e posizione *start* e l'apertura/chiusura di un marcatore vuoto con il medesimo *id* e posizione *end*. Di conseguenza, le due sequenze “a b c d e f” e “d e f g h i” possono essere trattate, per elaborazioni successive, *come se* vi fosse *overlapping* fra i marcatori. Anche con i fogli di stile è possibile ottenere lo stesso risultato, ma le strategie da applicare sono di solito più macchinose.

Vediamo ora rapidamente come si articola l'elaborazione dei repertori di congetture. In primo luogo, il programma estrae i nomi dei responsabili delle congetture (marcati con `...`) e li trasferisce in un database. Una volta raggruppati i dati, si procede manualmente a conformare i nomi abbreviati o scritti con differenti ortografie con l'*Index criticorum* dell'edizione di West, il quale per la sua sinteticità e completezza costituisce un buon sistema di riferimento.

Il programma estrae poi le parole delle congetture. Poiché l'attributo *n* delle congetture riporta la posizione della prima parola, il programma usa questo dato per assegnare la posizione corretta a ciascuna parola, inserendo nel database un numero progressivo oppure il numero indicato nell'attributo *n* della parola stessa e riprendendo poi la numerazione automatica. In questa fase, il programma confronta la posizione della parola e la parola stessa con i valori relativi all'edizione di riferimento. Per il repertorio costituito dall'apparato critico di Murray,

¹⁵Il marcatore `<anchor />` visto sopra è un esempio di *milestone*, in quanto rappresenta una posizione puntuale nel testo senza racchiudere nel corpo del marcatore una porzione di testo.

¹⁶Sull'argomento si veda anche SCHMIDT E COLOMB (2009).

3. Analisi di apparati e repertori

vengono registrate tutte le parole, per gli altri repertori, invece, qualora la forma presa in esame corrisponda a una forma dell'edizione di riferimento, nella medesima posizione, tale occorrenza viene contrassegnata come mero segnaposto. Si prenda di nuovo l'esempio della congettura di Blaydes: *διὰ χερὸς σφισὶν πρέπει*. Considerando il verso di appartenenza, cioè il v. 239 dell'edizione di Murray:

πότερα γὰρ τοξουλκὸς αἰχμὴ διὰ χεροῖν αὐτοῖς πρέπει

si vede che *διὰ* e *πρέπει* non hanno motivo di essere registrate come parole che devono figurare nell'indice alfabetico delle forme appartenenti alle congetture proposte da Blaydes. Merita attenzione, invece, *χερὸς*, dal momento che tale forma figura sì quale componente della congettura di Blaydes, ma in realtà appartiene alla paradosi, contro *χεροῖν*, che è congettura di Elmsley accolta da Murray, come si evince dagli apparati critici di Murray e di West presi in esame e marcati per il presente lavoro. In questo caso, la notizia nell'indice alfabetico delle forme presenti nei repertori di congetture ha il seguente aspetto:

χερὸς, 239: codd., Blaydes.

Negli indici e nelle concordanze delle congetture possono figurare quindi, quando siano deducibili dai repertori marcati, anche le difese esplicite della paradosi, quando siano contro l'edizione di riferimento.

Infine, il programma ricostruisce il contesto nel quale collocare la congettura allo scopo di creare le concordanze. È possibile scegliere se il contesto debba essere limitato al verso, oppure a un numero determinato di parole prima e dopo la forma presa in esame, oppure al periodo sintattico, limitato da due segni di interpunzione. Per la pubblicazione in rete possono essere valide la prima e la terza soluzione, per la destinazione a stampa sembra migliore la seconda.

Il procedimento seguito dal programma per ricostruire il contesto combina più fonti d'informazione, in quanto deve disporre, nelle posizioni appropriate, le parole appartenenti alla congettura (con relativa eventuale colometria e punteggiatura), le parole appartenenti all'edizione di riferimento del repertorio (Wecklein o West), quando non siano in accordo con l'edizione di Murray e infine le parole appartenenti all'edizione di riferimento generale, cioè Murray, sempre con relativa

3.3 Elaborazione dei documenti marcati

colometria e punteggiatura. Un esempio dovrebbe rendere più chiaro l'effettivo procedimento. Si prenda l'apparato di Wecklein al v. 50 (v. 49 Murray):

50. *σούνται* Blomfield. *πελάτης* Weil.

dopo la marcatura:

```

1 <app loc="50">
2 <span type="app_loc">50.</span>
3 <rdgGrp>
4 <rdg n="200" resp="Blomfield">
5 <w>σούνται</w>
6 <span type="rdg_resp">Blomfield</span>
7 <span type="rdg_grp_sep">.</span>
8 </rdg>
9 </rdgGrp>
10 <rdgGrp>
11 <rdg n="204" resp="Weil">
12 <w>πελάτης</w>
13 <span type="rdg_resp">Weil</span>
14 <span type="rdg_grp_sep">.</span>
15 </rdg>
16 </rdgGrp>
17 </app>

```

Murray ha accolto nel verso la congettura di Weil: *στεύται δ' ἱεροῦ Τμώλου πελάτης*, mentre l'edizione di Wecklein, seguendo l'indicazione dello sch. **M** (*οὕτως στεύται ἐνικὸν ἀντὶ πληθυντικοῦ*), concorda il verbo al singolare con il sostantivo al plurale della paradosi: *στεύται δ' ἱεροῦ Τμώλου πελάται*. Evidentemente, il programma deve collocare la congettura di Blomfield nel contesto dell'edizione di Wecklein, sulla quale si basa il repertorio che registra la congettura, e deve quindi combinare le informazioni in questo modo:

	200	2001	2002	2003	204
ed. di Murray	<i>στεύται</i>	<i>δ'</i>	<i>ἱεροῦ</i>	<i>Τμώλου</i>	<i>πελάτης</i>
ed. di Wecklein					<i>πελάται</i>
cong. di Blomf.	<i>σούνται</i>				
Blomf. in contesto	<i>σούνται</i>	<i>δ'</i>	<i>ἱεροῦ</i>	<i>Τμώλου</i>	<i>πελάται</i>

Per finire, il programma confronta automaticamente le congetture in contesto fra di loro, per verificare se alcune di esse sono ripetute in modo identico in repertori differenti¹⁷, o se le stesse congetture sono state proposte da autori diversi,

¹⁷Quasi tutte le congetture che si trovano nell'apparato di West sono, per motivi comprensibili, già registrate nei repertori di Wecklein o di Dawe.

3. Analisi di apparati e repertori

senza un'opportuna segnalazione da parte dei curatori dei repertori. Si interviene quindi manualmente per marcare nel database quali informazioni escludere dalla redazione degli indici e delle concordanze.

Elaborazione e stampa dei risultati

I dati contenuti nel database vengono poi elaborati da un secondo programma che produce automaticamente dei file XHTML per la pubblicazione dei risultati in rete oppure dei file XML associati ai relativi fogli di stile XSL-FO¹⁸, per la pubblicazione nel formato PDF, adatto sia alla consultazione a video sia alla stampa. Tanto gli indici quanto le concordanze prodotti possono essere ordinati seguendo l'ordine alfabetico delle forme, oppure l'ordine alfabetico dei responsabili delle congetture. Questa seconda modalità si rivela utile soprattutto in vista di un arricchimento delle informazioni contenute nei repertori attraverso l'esame effettivo dei testi che costituiscono la fonte dei curatori dei repertori.

A puro titolo di esempio¹⁹, per illustrare uno dei possibili impieghi del database oltre alla creazione di indici e concordanze delle congetture nel contesto del verso di appartenenza, si mostra di seguito una tabella di nomi propri sui quali siano state formulate congetture. Si vuole mostrare come sia possibile filtrare i dati in base ad un criterio linguistico, codificato nella base di dati di riferimento, al fine di individuare, nella base di dati delle congetture, le informazioni corrispondenti.

<i>Ἄγαβάτας</i>	960	<i>Ἄγαβάτας</i>	<i>Ἄβαγάθας</i> Roussel – <i>Ἄβαγτάνας</i> Hannak – <i>Ἄγδαβάτας</i> C. Schneider
<i>Ἄγδαδάτης</i>	997	<i>κῆγδαδάταν</i>	<i>Γαδάταν</i> Weil ¹¹ – <i>Δαδάκαν</i> Wil. ¹⁹ – <i>Δαδάταν</i> Weil ¹¹ – <i>κᾶγδαδάταν</i> Keiper – <i>Κυγδαγάταν</i> Passow
<i>Ἄγχάρης</i>	994	<i>Ἄγχάρην</i>	<i>ἀγχάρη</i> Humbach
<i>Ἄδεύης</i>	312	<i>Ἄδεύης</i>	<i>ἀδειῆς</i> van Hoff.
<i>Ἄιδωνεύς</i>	649	<i>Ἄιδωνεύς</i>	<i>Ἄιδωνεῦ</i> Blomf.

¹⁸Si veda, per una panoramica su questa tecnologia, il sito: <http://xml.apache.org/fop>.

¹⁹Volutamente l'elenco non è né esaustivo né controllato a mano.

3.3 Elaborazione dei documenti marcati

	650	Ἰδωνεύς	Ἰδωνεὺ Blomf.
ἸΑλπιστος	981	ἸΑλπιστον	ἸΑλπίσταν Heimsoeth – ἄλπιστον Wackernagel ³ – ἄλπνιστον Wecklein ¹¹
ἸΑμιστρις	320	ἸΑμιστρις	ἸΑμηστρις Brunck ¹
ἸΑραβος	318	ἸΑραβος	ἸΑραμβος Weil ¹⁵
ἸΑριόμαρδος	321	ἸΑριόμαρδος	ἸΑριομᾶρ – ἸΑριόμαρδε Weil
	968	καριόμαρδος	Καριόμαρδός Asul.
ἸΑρίων	994	ἸΑρίων	ἄρειον Roussel – ἀρείονα Heimsoeth – ἸΑρίον H.L. Ahrens ¹ 10n.
ἸΑρκεύς	312	ἸΑρκεύς	ἸΑρτεὺς Hartung
ἸΑρσάκης	995	ἸΑρσάκης	ἸΑρσάμην Burney ¹ – ἸΑστάσπην Hartung
ἸΑρσάμης	308	ἸΑρσάμης	ἸΑρσάκης Hartung
ἸΑρτάβης	318	ἸΑρτάβης	ἸΑρσάκης Hartung
Βατάνωχος	981	Βατανώχου	Βαγανώχου R. Schmitt 54
Βόλβη	494	Βόλβης	Βοίβης Casaub.
Δαριείαν	651	Δαριάνα	Δαριαῖον Wecklein – Δαριαίων Roussel – δ' ἄρει' ἄν G.C.W. Schneider – ἄνα Δαρεῖ' Haupt – δ' αἶρε ... Δαριείαν Rossbach – Δα- ριανῆ Weil – δαίμονα Newman – δάϊον Herm. – Δαρεῖ' ἄνα Siebelis – Δαριεῖα Lachmann – Δαρεῖον ἄνακτα Bothe – ἸΕρξείαν Blomf. – Δαριανεὺ Weil ¹¹
	663	Δαριάν	
Δίαιξις	671	Δίαιξιν	Διάϊξιν G.C.W. Schneider – Διά- ναιξιν Heimsoeth – Δαῖξιν Har- tung
Δοτάμας	959	Δοτάμας	Δαδάκας Heimsoeth – Δατάμας Passow ¹ 52 – Δότμας Hartung
Μάρδος	774	Μάρδος	Μᾶγος Lange-Pinzger – Μέρδης Rutgers 231 – Σμέρδης Blomf.
Μάρδων	51	Μάρδων	Μάνδρων Weil ¹¹
Μασίστρης	30	Μασίστρης	Μασίστης Weil ¹¹
	971	Μασίστρας	Μασίστας Weil ¹¹
Μάταλλος	314	Μάταλλος	Μάταλος Bothe ¹

3. Analisi di apparati e repertori

<i>Μεγαβάτης</i>	22	<i>Μεγαβάτης</i>	<i>Μεγαβάταν</i> Rob.
<i>Μητρογαθής</i>	43	<i>Μητρογαθής</i>	<i>Μιθραγαθής</i> Blomf. – <i>Μιτραγαθής</i> C.G. Schütz
<i>Ξάνθης</i>	994	<i>Ξάνθην</i>	<i>Ξάνθιν</i> G.C.W. Schneider – <i>Ξάνθον</i> Roussel
<i>Ξέρξης</i>	923	<i>Ξέρξα</i>	<i>Ἄρει</i> Hartung – <i>Ξέρξα</i> Pauw
<i>Οιβάρης</i>	983	<i>Οιβάρην</i>	<i>Οιβάζην</i> Blomf. – <i>Ἰοιβάρην</i> Heimsoeth – <i>Οιβάρην</i> West
<i>Πάρθος</i>	983	<i>Πάρθον</i>	<i>Μάρδων</i> Haupt – <i>φθαρτὸν</i> Heimsoeth
<i>Πελάγων</i>	959	<i>Πελάγων</i>	<i>Πήγας Τενάγων</i> Hartung
<i>Πέρσης</i>	978	<i>Πέρσαν</i>	<i>Περσῶν</i> Roussel – <i>Πέρσην</i> Blomf. – <i>Περσῶν</i> G.H. Schütz
<i>Πηγασταγών</i>	35	<i>Πηγασταγών</i>	<i>ἐπὶ γᾶς ταγών</i> Rogers – <i>Πήγας</i> Blomf. ⁶ – <i>πέλτης ταγών</i> Siebelis – <i>πηγαῖς ταγών</i> Teuffel – <i>πηγᾶς ταγών, ταγὸς πηγῶν</i> Prien – <i>Πελαγών</i> Hartung
<i>Σευάλκης</i>	969	<i>Σευάλκης</i>	<i>Σευάκης</i> Asul.
<i>Σήσαμον</i>	982	<i>Σησάμα</i>	<i>Σεισάμαν</i> Heimsoeth – <i>Σησάμαν</i> Blomf. – <i>Σησαμάτου</i> Rob. – <i>Σισάμνα</i> L. Dindorf
<i>Σοσθάνης</i>	32	<i>Σοσθάνης</i>	<i>Ἰοσθάνης</i> Kiesser – <i>Σοοσθάνης</i> Burges – <i>Σωσθάνης</i> Blomf.
<i>Τενάγων</i>	306	<i>Τενάγων</i>	<i>Πελάγων</i> Hartung
<i>Τόλμος</i>	998	<i>Τόλμον</i>	<i>Τόλμαν</i> V.D. ^{Dw}
<i>Ἰσταίχμας</i>	972	<i>Ἰσταίχμας</i>	<i>Ἰσταίχμης</i> Sanneg
<i>Φαρανδάκης</i>	31	<i>Φαρανδάκης</i>	<i>Φαρανδάτης</i> Oberdick
<i>Φαρνοῦχος</i>	313	<i>Φαρνοῦχος</i>	<i>Ἄρνοῦχος</i> Pauw – <i>Καρδοῦχος</i> Weil – <i>Φαροῦχος</i> G.C.W. Schneider – <i>Φερνοῦξος</i> Rogers – <i>ῥαβδοῦχος</i> Weil
<i>Χρυσεύς</i>	314	<i>Χρυσεύς</i>	<i>Χρυσοῦς</i> Bothe ¹
<i>Ψάμμης</i>	959	<i>Ψάμμης</i>	<i>Ψαμμισκάνης</i> Weil ¹¹

4

Trattamento automatico di apparati e repertori

Lo studio svolto nel capitolo precedente su campioni testuali di apparati critici e repertori di congetture permette di formalizzare le procedure per il trattamento automatico di documenti che abbiano la stessa struttura.

4.1 Stato dell'arte

Attualmente, ci sono due approcci principali per aggiungere l'apparato critico alle edizioni digitali. Il primo è basato sulla collazione automatica di edizioni diplomatiche. Le edizioni diplomatiche digitali sono trascrizioni complete di singoli manoscritti, arricchiti di informazioni relative alla *mise en page*, alla posizione e alla funzione di ciascuna porzione del testo nella pagina, etc. Usualmente sono codificate in XML, secondo le linee-guida della Text Encoding Initiative¹. Possono essere usate per stampare il testo del testimone, per mappare (e cercare) il testo digitale sull'immagine della pagina o per collazioni automatiche. Tramite il linguaggio di marcatura offerto dalla TEI è possibile separare il testo del manoscritto dalle sue interpretazioni: correzioni, normalizzazioni, scioglimento di abbreviazioni, etc. Questo metodo è particolarmente utile con un piccolo nu-

¹<http://www.tei-c.org>

4. Trattamento automatico di apparati e repertori

mero di manoscritti e in assenza di vasta letteratura secondaria costituita da commentari e articoli.

Il secondo approccio è basato sull'inserimento manuale di notizie d'apparato tramite piattaforme collaborative online. È utile se lo scopo è l'acquisizione di ampie quantità di informazioni d'apparato, su molti testi di vari autori o su tradizioni complesse. Questo metodo è stato usato ad esempio nel progetto *Musisque Deoque* con lo scopo di fornire, per un numero molto elevato di poeti latini, almeno un apparato critico minimo. In effetti, è meglio avere informazioni d'apparato essenziali per un *corpus* molto ampio anziché avere un apparato molto dettagliato per un numero ristretto di opere e di autori. Per lo studio della complessa tradizione di Avicenna, il progetto *Philosophy on the Border of Civilizations*² sta sviluppando l'editor CEED, un editor per edizioni critiche dove le varianti vengono inserite manualmente attraverso una piattaforma web.

L'allineamento completo della base di collazione (edizione di riferimento o *codex optimus*) con le diplomatiche digitali di ciascun testimone è molto utile per lo studio delle relazioni fra manoscritti, ma non è applicabile in tutte le situazioni. Quando si abbia una tradizione diretta e indiretta molto ampia e una ricca letteratura secondaria, non è sempre possibile ricostruire un contesto per la variante o la congettura che sia ampio quanto l'intero testo. La variante che viene estratta da uno *scholium* ha un contesto indefinito, perché non si può sapere quale fosse esattamente l'intero testo letto dallo scoliasta. Inoltre, le congetture spesso sono espresse in forma disgiuntiva: a *vel* b *vel* c, e a volte non si può sapere quale sia stata l'edizione di riferimento usata dallo studioso che ha formulato la congettura. Se le edizioni diplomatiche sono simili a strati che si possono sovrapporre, questi casi sono invece frammenti d'informazione che non si sa con precisione dove debbano essere collocati. Se n edizioni diplomatiche possono essere distribuite su n dimensioni, questi frammenti d'informazione con un contesto indefinito teoricamente esistono in topologie più complesse, che generano un numero incontrollabile di combinazioni.

²<http://www.avicennaproject.eu>

4.2 Edizioni critiche di riferimento e repertori di congetture

La metodologia dell’inserimento manuale delle notizie d’apparato ha un limite nella soggettività degli studiosi che inseriscono i dati. Devono infatti decidere come interpretare l’informazione originale dell’apparato a stampa ai campi del modulo, come integrare eventuali informazioni mancanti e come omettere le informazioni di scarso rilievo. Inoltre, l’informazione estratta e rielaborata non è più direttamente mappabile sull’apparato originario. La TEI fornisce indicazioni per questo tipo di mappatura, ma la marcatura manuale è molto dispendiosa, in termini di tempo, per grandi quantità di testo. Un compromesso accettabile è dato dal *parsing* automatico di apparati e repertori, congiuntamente alla collazione automatica di un gruppo di trascrizioni diplomatiche di rilievo. Le scelte soggettive degli studiosi sono limitate alle fasi di correzione. Questo terzo approccio ha un duplice scopo: da un lato è finalizzato ad analizzare automaticamente apparati critici e repertori di congetture su Eschilo e dall’altro è finalizzato a individuare euristiche utili per qualsiasi collezione di varianti e/o congetture con una struttura simile. La mappatura accurata di informazioni estratte da apparati e repertori deve essere usata per costruire nuove edizioni critiche digitali, indici, concordanze e sistemi per il *text retrieval* basato su varianti³.

4.2 Edizioni critiche di riferimento e repertori di congetture

Il primo problema da affrontare è l’edizione di riferimento, cioè il testo che costituisce la base per indici e concordanze, il riferimento per i commenti e la letteratura secondaria, il sistema di numerazione dei versi per gli apparati e i repertori. Di solito l’edizione di riferimento è l’edizione più autorevole per la comunità scientifica corrente. Tuttavia, quando una nuova edizione autorevole

³Per seguire il dibattito sul *Digital Scholarly Editing*, sui modelli per rappresentare le varianti, sull’*information retrieval* applicato alla variantistica, si vedano fra gli altri SCHMIDT E COLOMB (2009), VANHOUTTE (2010), CIOTTI (2011), PIERAZZO (2011), ALMAS E BEAULIEU (2013), SAHLE (2013), APOLLON E BELISLE (2014), PIERAZZO (2015), DRISCOLL E PIERAZZO (2016).

4. Trattamento automatico di apparati e repertori

sostituisce la precedente, i vecchi e i nuovi strumenti di sussidio, come indici e concordanze, si trovano a essere basati su testi diversi. Nello specifico, il presente lavoro su Eschilo usa tre differenti edizioni di riferimento, perché l'apparato critico dei repertori delle congetture di WECKLEIN (1885) e WECKLEIN (1893) sono basati su WECKLEIN (1885) stesso, le collazioni di manoscritti di DAWE (1963) e il suo repertorio di congetture DAWE (1965) sono basati su MURRAY (1955), mentre l'appendice delle congetture raccolte da WEST (1990) e il suo stesso apparato sono basati su WEST (1998). Un'edizione può differire da un'altra non solo per le varianti testuali, ma anche per la disposizione dei versi, distribuiti in maniera differente su più linee di testo, secondo le interpretazioni metriche e colometriche proposte dall'editore. In questo modo, il riferimento al numero del verso non è uno strumento efficace per passare da un'edizione di riferimento a un'altra, perché è ambiguo: ad esempio i versi di *Pers.* 857-8 (WECKLEIN, 1885) *πανταρκής, ἀκάκας, | ἄμαχος βασιλεύς* non hanno la stessa distribuzione sui versi 855-56 (MURRAY, 1955) *πανταρκής ἀκάκας ἄμαχος βασι-|λεύς ...* a causa di una differente colometria. Solo la posizione sequenziale delle parole nell'intero testo fornisce la matrice di riferimento per passare da un'edizione all'altra, e anche la numerazione della colometria e dei versi si basa su questa matrice: ad esempio *βασιλεύς* è nella 4429ma posizione testuale in entrambe le edizioni, ma la nuova riga è mappata sull'ultimo carattere della parola in WECKLEIN (1885) e sul quarto carattere in MURRAY (1955).

Apparati e repertori, redatti nel corso dell'Ottocento e del Novecento, si differenziano per convenzioni tipografiche e quantità di informazioni, più o meno accurate. Tuttavia, l'assunto di base è che è possibile identificare un piccolo numero di schemi ed espressioni ampiamente ripetuti, al fine di marcare automaticamente ogni informazione analizzata.

La struttura più semplice è data da *congettura editore*, dove la congettura è espressa da una più parole greche. Esistono tuttavia casi più complessi. Ad esempio, nel repertorio di Wecklein, i numeri di verso sono seguiti dal segno di punteggiatura e le diverse edizioni dello stesso editore sono indicate dalla formula: *congettura*₁ olim, postea *congettura*₂ editore, come si vede nel seguente esempio:

4.3 Formalizzazione della struttura di apparati e congetture

132. λέκτρα δ' ἀμῖν μάταν Enger. πόνω Pauw, σπάνει olim, postea ἔρω Heimsoeth, ὄδω Oberdick. Nel repertorio di Dawe, i numeri di verso non sono seguiti dal segno di punteggiatura e le citazioni, relative alle riviste e alle monografie, sono seguite dal numero di pagina, come si vede nel seguente esempio:

133 λέκτρα δ' ἀντ' ἀνέρων πόθωι Hoernle p. 89

4.3 Formalizzazione della struttura di apparati e congetture

Le operazioni testuali registrate in apparati critici e repertori di congetture possono essere ricondotte a quattro tipi: inserzioni, cancellazioni, sostituzioni e trasposizioni. Anche se le inserzioni possono assumere la funzione specifica di ripetizione e le cancellazioni possono essere *lacunae* o omissioni, da un punto di vista computazionale le operazioni di base sono le stesse e consentono qualsiasi trasformazione dalla stringa di origine alla stringa di destinazione. La sostituzione può essere ricondotta a una cancellazione di testo e a un'inserzione di un testo diverso nella stessa posizione, mentre la trasposizione può essere ricondotta a una cancellazione seguita da un'inserzione in un'altra posizione dello stesso testo.

operazione	stato iniziale	stato finale
inserzione	a c	a b c
cancellazione	a b c	a c
sostituzione	a	b
trasposizione	a b, a b c	b a, c b a

Secondo statistiche eseguite su un campione del 5% del testo di Eschilo, negli apparati e nei repertori presi in esame circa il 90% di varianti e congetture sono espresse solo da numero di verso e da sequenze di parole greche, seguite da liste di testimoni (in caso di varianti) o di nomi di studiosi (in caso di congetture). Nella maggior parte dei casi la sequenza di parole greche rappresenta una semplice sostituzione testuale, ma a volte l'informazione è costituita da segnaposto (parole

4. Trattamento automatico di apparati e repertori

di confine identiche a quelle presenti nell'edizione di riferimento) che servono a indicare la posizione corretta per ancorare una lezione contenente un'aggiunta, una cancellazione o una trasposizione di testo. Ad esempio, dato il verso:

370 ναυσὶν κρυφαίως δρασμὸν εὐρόντες τινά,

nel repertorio di Wecklein si trova:

370. δρασμὸν ἄραντες Naber.

Il rimanente 10% di varianti e congetture è composto da strutture più complesse, con una frase latina che esprime l'operazione testuale che deve essere eseguita (*delet, iterat, transponit, etc.*). Ad esempio, dato il verso:

3 καὶ τῶν ἀφνεῶν καὶ πολυχρύσων ἐδράνων φύλακες

nel repertorio di Wecklein si trova:

3. καὶ πολυχρύσων delet Bothe.

Il caso più semplice (e fortunatamente piuttosto frequente) si verifica quando la lezione è una variante ortografica o morfologica che sostituisce una singola parola nell'edizione di riferimento. Qualche volta però la variante divide la parola in due parti: ad esempio ἐν τλήμονι invece di εὐτλήμονι. Quando la sostituzione è una glossa, un sinonimo, un iperonimo, un iponimo o una parola non correlata in alcun modo, nell'apparato e nei repertori può essere indicato dalla formula x: y o x *pro* y (ad esempio κίοντων Wecklein : ἰόντων codd.). Quando la sostituzione è ampia e complessa e contiene eventuali cancellazioni o aggiunte di testo, usualmente la prima e l'ultima parola sono dei segnaposto per ancorare la lezione all'edizione di riferimento.

La cancellazione di solito è indicata dalle parole da cancellare, seguite dall'espressione *delet* (ad esempio: καὶ πολυχρύσων delet Bothe). L'inserzione di parole usualmente è indicata dalla formula [*ante/post* x] *addit* y, dove x è la parola dell'edizione di riferimento (ad esempio: ante βαλλήν *addit* ἰωὰ Dindorf). La trasposizione è la combinazione di una cancellazione e di un'inserzione di testo. Può essere una semplice inversione di parole o può interessare uno o più versi (ad esempio: 94-102 post 116 *transponit* OMueller). La fonte è uno o più manoscritti per le varianti o uno o più studiosi per le congetture, che possono essere seguite da

4.3 Formalizzazione della struttura di apparati e congetture

un'accurata indicazione bibliografica. Apparati e repertori diversi possono usare differenti abbreviazioni per i nomi dei manoscritti e degli studiosi. L'informazione sulle fonti può avere diversi gradi di precisione. Ad esempio, nell'apparato di West ciascun manoscritto è sempre identificato, mentre nel repertorio di Wecklein i manoscritti diversi da **M** (il *codex optimus*) sono contrassegnati con recc. (*recentiores*). Nell'apparato di West ciascuna edizione moderna è identificata dal nome del suo autore con un numero in apice (ad esempio: Bothe³), mentre nel repertorio di Wecklein le edizioni precedenti sono distinte dall'ultima tramite l'espressione *olim* x (ad esempio: olim Bothe).

Lo studio delle strutture di apparati e repertori, facilitato dall'annotazione manuale di campioni, ha portato alla definizione di un Domain Specific Language (DSL), interpretato da ANTLR⁴, per classificare automaticamente tutti gli elementi delle notizie di apparato (numero di verso, sequenza di parole greche, parole latine, *sigla* di manoscritti o nomi di studiosi e riferimenti bibliografici).

Come si può vedere in Fig. 4.1, tramite ANTLR è possibile scrivere le regole necessarie a decodificare, e quindi a classificare, le informazioni dei repertori. La scrittura delle regole va dal generale al particolare (*top-down*), per cui prima si definisce il contenuto del repertorio nel suo insieme come lista di notizie d'apparato; poi si definisce la singola notizia d'apparato come una sequenza costituita da numero di verso, uno o più gruppi di lezioni e a capo; poi si definiscono il gruppo e la singola lezione, fino a raggiungere gli elementi costitutivi minimi. È da notare che ciò che si vede in figura è solo una piccola parte, semplificata, dell'intera grammatica per decodificare il contenuto dei repertori.

Applicando ANTLR al repertorio di Wecklein, si possono quindi classificare le parti costituenti delle varie congetture, in modo da estrarre le informazioni senza necessità di marcatura manuale. Ad esempio, date le seguenti linee del repertorio:

325 *εὐθέτως* Weil. 326 *δὲ* Paley. 327 *ἀπ' ἀρχῆς* Pauw, *ὑπαρχος* Hermann,
τάγαρχος Meineke. *πλείστον εἰς ἀνὴρ* Dindorf.

⁴Si veda PARR (2014). Il software è disponibile al sito <http://antlr.org>

4. Trattamento automatico di apparati e repertori

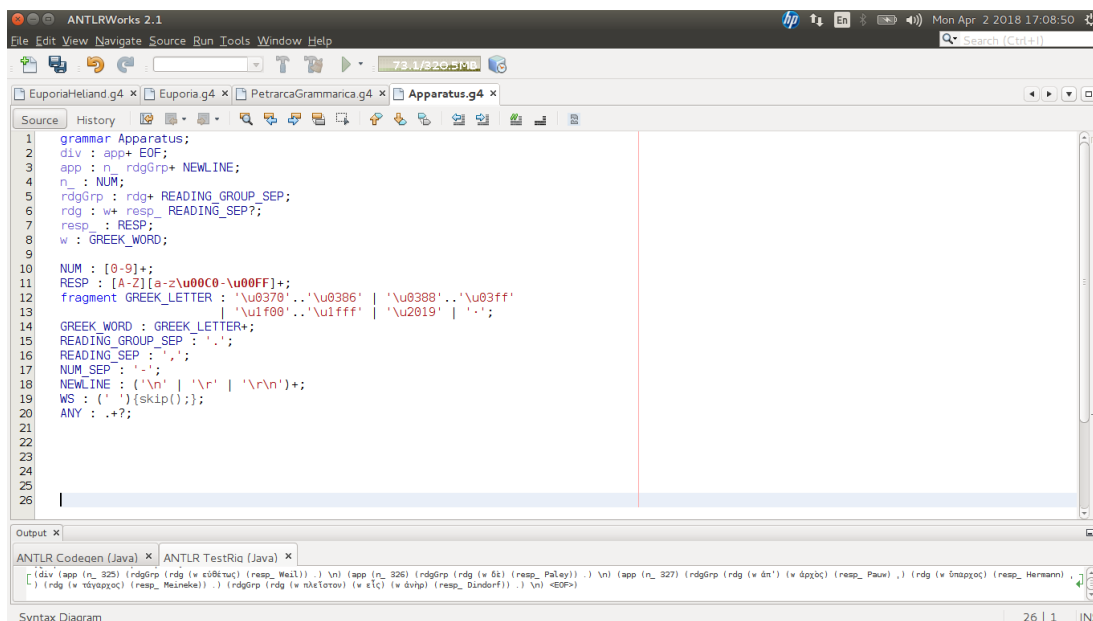


Figura 4.1: Regole per decodificare le informazioni dei repertori

La Fig. 4.2 mostra come vengono classificate le informazioni per poter essere estratte, marcate automaticamente o inserite subito in un database in forma strutturata senza passare attraverso la marcatura.

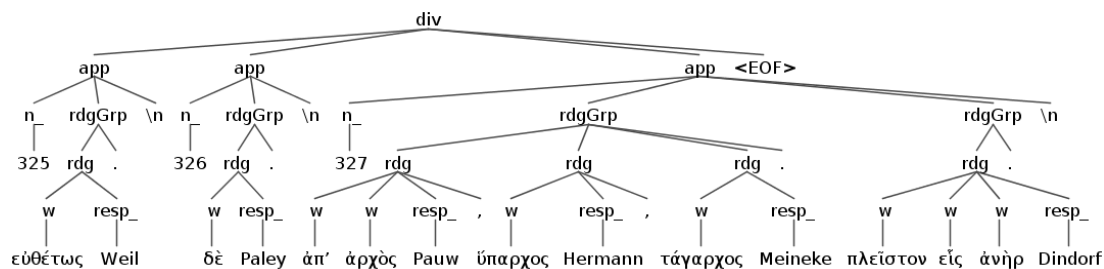


Figura 4.2: Estrazione di informazioni dal repertorio di Wecklein con ANTLR

4.4 Metodi di allineamento

Come visto sopra, circa il 90% delle lezioni è costituito dall'indicazione del verso, da una sequenza di parole greche alternative alla lezione dell'edizione di riferimento e dall'indicazione del manoscritto che reca la variante o dall'indicazione dello studioso che ha proposto la congettura. La variante o la congettura viene mappata sul testo dell'edizione di riferimento grazie a un algoritmo di allineamento⁵.

In realtà, conoscere in quale verso eseguire la sostituzione non è sufficiente; si deve conoscere la posizione precisa di ciascuna parola all'interno del verso per creare indici e concordanze di varianti e congetture. Una concordanza deve ricostruire il contesto locale della parola e i sistemi di *information retrieval*, quando eseguono ricerche su più parole, devono riconoscere le parole in posizioni adiacenti.

Negli apparati e nei repertori, le varianti e le congetture sono localizzate solo attraverso il riferimento al verso, non alla posizione precisa nel verso. Questa informazione è superflua per gli studiosi, ma non per il calcolatore. Gli algoritmi di allineamento, che valutano la similarità di una stringa con un'altra stringa (o sotto-stringa) sono basati sull'*edit distance*⁶, cioè sulla valutazione dei costi per eseguire inserzioni, sottrazioni e sostituzioni per trasformare la prima stringa nella seconda o parte di essa. Seguendo questo principio, ogni lezione può essere allineata con la porzione di testo nell'edizione di riferimento con l'*edit distance* minore (e dunque con la similarità maggiore).

L'allineamento delle lezioni con l'edizione di riferimento può essere eseguito a diversi livelli di granularità, a seconda degli scopi.

Un allineamento sequenza per sequenza identifica la parte del verso (o dei versi) nell'edizione di riferimento dove la variante (o la congettura) dovrebbe andare a collocarsi.

Questo tipo di allineamento è adatto per *corpora* non annotati, per *corpora* annotati a livello di verso o di unità ancora più ampie. L'allineamento sequenza

⁵Sull'allineamento finalizzato allo studio di testi storici si veda BÜCHLER ET AL. (2014) e YOUSEF E PALLADINO (2017).

⁶Si veda ad esempio MARZAL E VIDAL (1993).

4. Trattamento automatico di apparati e repertori

Ed. di rif.	<i>Νείλος ἔπεμψεν</i>	<i>Σουσισκάνης, </i>	<i>Πηγαστάγων</i>	<i>Αἴγυπτογενής</i>
Blomfield	<i>Νείλος ἔπεμψεν</i>	<i>Σουσας, Κάνης,</i>	<i>Πήγας, Πελάγων</i>	<i>Αἴγυπτογενής</i>

Tabella 4.1: Allineamento sequenza per sequenza (*Pers.* 35-36)

per sequenza è preferibile nei casi di collegamenti eseguiti manualmente, perché basta determinare il punto iniziale e finale, riducendo le scelte individuali. Questa soluzione è stata adottata ad esempio nel progetto *Musisque Deoque*.

Quando i *corpora* sono arricchiti di varianti e congetture, è opportuno che l'informazione ridondante o irrilevante sia ignorata, escludendo le parole della lezione che hanno la mera funzione di segnaposto. La tabella 4.2 mostra i modi possibili per mappare la congettura *οὐδαμ' οὖσ' ἐμαυτῆς* L. Schmidt su *Pers.* 165 *μῦθον οὐδαμῶς ἐμαυτῆς οὖσ' ἀδείμαντος, φίλοι*. L'ultima parola, *ἐμαυτῆς*, aiuta il lettore a trovare la posizione corretta della congettura: la fissa nel contesto dell'edizione di riferimento, ma non è una componente necessaria della lezione. Se gli

Allineamento sequenza per sequenza

Ediz. di rif.	<i>μῦθον</i>	<i>οὐδαμῶς ἐμαυτῆς</i>	<i>οὖσ' ἀδείμαντος, φίλοι</i>
L. Schmidt		<i>οὐδαμ' οὖσ' ἐμαυτῆς</i>	

Allineamento parola per parola e rimozione dei segnaposto

Ediz. di rif.	<i>μῦθον</i>	<i>οὐδαμῶς</i>	–	<i>ἐμαυτῆς οὖσ' ἀδείμαντος, φίλοι</i>
L. Schmidt		<i>οὐδαμ'</i>	<i>οὖσ'</i>	<i>ἐμαυτῆς</i>

Tabella 4.2: Identificazione dei segnaposto (*Pers.* 165)

elementi dell'edizione di riferimento sono annotati linguisticamente, metricamente o stilisticamente, anche le lezioni estratte dai repertori lo dovrebbero essere, seguendo i medesimi criteri. Spesso le lezioni (o loro parti) hanno in comune tratti linguistici o metrici (stesso lemma, stessa parte del discorso, stessa struttura metrica). In questo caso, l'allineamento parola per parola permette di evidenziare i tratti comuni.

Ediz. di rif.	<i>βάσκε</i>	<i>πάτερ</i>	<i>ἄκακε</i>	–	<i>Δαριάν</i>	<i>οἶ.</i>
FWNewman	<i>βάσκε</i>	<i>πατήρ</i>	<i>ἀκάκας</i>	<i>ὁ</i>	<i>Περσῶν</i>	
Lemma comune		<i>πατήρ</i>	<i>ἀκάκας</i>			
Categoria grammaticale comune		Nome	Aggettivo		Nome	

Tabella 4.3: Caratteristiche condivise fra gli elementi allineati (*Pers.* 668)

Dopo correzioni manuali e integrazioni, l'allineamento parola per parola è utile per classificare elementi rilevanti della lezione, al fine di identificare varianti ortografiche (stesso lemma e stessi tratti morfologici), lessicali (lemmi diversi) e morfologici (stesso lemma ma differenti tratti morfologici).

L'allineamento carattere per carattere è adatto quando è possibile assegnare a ciascun manoscritto o a ciascuna edizione moderna un *layer* indipendente ed è particolarmente utile per lo studio di errori causati dalla *scriptio continua*. Con questo tipo di allineamento è possibile estrarre statistiche sulle sostituzioni di caratteri, per studi di tipo paleografico. I classici algoritmi di allineamento prendono in considerazione sostituzioni, inserzioni e cancellazioni, ma esistono versioni modificate che prendono in considerazione anche la trasposizione di segmenti adiacenti o la compressione e l'espansione, dove due o più unità di una stringa corrispondono a una singola unità dell'altra⁷.

Ediz. di rif.	<i>πλαγκτοῖς ἐν διπλάκεσσιν.</i>
	<i>ΠΛΑΓΚΤΟΙΣΕΝ - - ΔΙΠΛΑ - ΚΕΣΣΙΝ</i>
	<i>ΠΛΑΓΚΤ - - - ΕΝΣΠ - Ι - ΛΑΔ - ΕΣΣΙΝ</i>
Hartung	<i>πλάγκτ' ἐν σπιλάδεσσιν</i>

Tabella 4.4: Allineamento carattere per carattere (*Pers.* 280)

⁷KONDRAK (2002) spiega l'applicazione di questi algoritmi alla linguistica storica.

4. Trattamento automatico di apparati e repertori

In NAVARRO E RAFFINOT (2002) e CROCHEMORE ET AL. (2007) si trova la spiegazione dettagliata dell'allineamento locale (Smith-Waterman) e dell'allineamento globale (Needleman-Wunsch). Mentre l'allineamento locale individua regioni simili, nel modo più compatto possibile, in sequenze di lunghezza diversa. l'allineamento globale invece individua somiglianze fra sequenze di lunghezza simile.

L'allineamento del contenuto della congettura con la corretta porzione di verso dell'edizione di riferimento avviene in due fasi. Nella prima fase, si identificano le estremità del segmento di verso interessato, valutando l'*edit distance* minore fra ciascuna parola del verso e ciascuna parola della congettura e prendendo, a parità di valore, l'estremo sinistro e l'estremo destro. In questo modo è possibile identificare la porzione di testo interessata, anche in caso di inversione di parole o altri tipi di breve trasposizione. Ad esempio, data la congettura *πλείστον εἰς ἀνὴρ* di Dindorf al v. 327 *Κιλικῶν ἄπαρχος, εἰς ἀνὴρ πλείστον πόνον*, la sequenza di parole *πλείστον εἰς ἀνὴρ* viene facilmente allineata a *εἰς ἀνὴρ πλείστον*. Se non si tiene conto delle possibili trasposizioni, anche un algoritmo di allineamento locale può essere utile per individuare la posizione della congettura nel verso, e inoltre i due metodi possono essere usati insieme per posizionare correttamente sia congetture come quella appena vista, sia congetture come *ἀπ' ἀρχὸς* di Pauw, rispetto ad *ἄπαρχος*, nello stesso verso. Nella seconda fase si applica l'algoritmo di allineamento globale per far corrispondere correttamente ciascuna parola (o ciascun carattere) della congettura a una parola (o a un carattere) del segmento individuato nella prima fase, oppure a un *gap*.

5

Modelli, risorse e strumenti per l'analisi sintattica

L'obiettivo principale di questo tipo di analisi nell'ambito della presente ricerca è la possibilità di confrontare strutture sintattiche simili o identiche presenti nel testo, di misurare il grado di complessità di un periodo, il numero di occorrenze di un determinato sintagma, etc.

Come si vedrà, la ricerca attuale nel campo dell'analisi sintattica con metodologie informatiche ha portato alla creazione di strumenti che hanno come obiettivo la descrizione di strutture sintattiche e non la creazione di repertori di interpretazioni sintattiche storicamente determinate. In ambito filologico è impossibile ignorare il fatto che differenti lezioni possono influire sulla sintassi del periodo a cui appartengono e che una medesima lezione può essere stata oggetto di differenti interpretazioni sintattiche, registrate in commentari e studi specifici. Per il presente studio, non essendo stato possibile procedere in modo sistematico ed esaustivo alla registrazione di tutte le interpretazioni alternative documentate, si è preferito procedere in due differenti direzioni. In primo luogo, si è condotta l'analisi sull'intero testo dei *Persiani* secondo l'edizione di Murray, concependo la descrizione sintattica come *interpretazione accolta*, le cui scelte sono giustificate dalla consultazione sistematica di KÜHNER E GERTH (1898), SCHWYZER (1990), ADRADOS (1992), MATINO (1998), dei commentari di GROENEBOOM

5. Analisi sintattica

(1930), BROADHEAD (1960), ROUSSEL (1960) e BELLONI (1994); inoltre dalla consultazione, per casi notevoli, dei commentari e della letteratura specifica citati in bibliografia generale. In secondo luogo, si è ideato un modello teorico per affrontare il problema di codificare la relazione molti-a-molti data da: una lezione / molte interpretazioni storicamente attestate; molte lezioni / una o più interpretazioni. Per verificare la bontà del modello, lo si è applicato ad alcuni casi di particolare interesse, di cui si parlerà nel commentario, al Cap. 8.

L'impiego del calcolatore

La bibliografia sulle metodologie informatiche per condurre l'analisi sintattica è piuttosto estesa: rimandiamo pertanto ai repertori bibliografici consultabili online della *Association for Computational Linguistics*¹ (ACL) e della *Association for Computing Machinery*² (ACM), limitandoci a segnalare due manuali introduttivi di buon livello, anche se datati: ALLEN (1995) e BOGURAEV E PUSTEJOVSKY (1996) e una rassegna aggiornata sulla creazione e l'uso di banche dati di strutture sintattiche: ABEILLÉ (2003), cui si farà più volte riferimento nel corso di questo capitolo.

Nel loro complesso, le metodologie adottate possono essere ricondotte a due grandi categorie: da un lato l'analisi dei costituenti, praticata in seno alla grammatica generativo-trasformazionale e dalle sue derivate, in particolare dalla grammatica lessico-funzionale; e dall'altro l'analisi della reggenza, praticata in seno alle grammatiche di dipendenza. Si cercherà quindi, molto brevemente, di riassumere i vantaggi e gli svantaggi delle formalizzazioni proposte dalle rispettive teorie.

L'analisi dei costituenti nella grammatica generativo-trasformativale

Per quanto la grammatica generativo-trasformativale vanta numerosi seguaci, tuttavia le tappe principali della sua evoluzione si possono trovare nelle principali

¹<<http://acl.ldc.upenn.edu>>

²<<http://www.acm.org>>

opere di CHOMSKY, da *Syntactic Structures* (1957), attraverso *Aspects of the Theory of Syntax* (1965), *Lectures on Government and Binding* (1984), fino a *The Minimalist Program* (1995), per menzionare soltanto i testi dove l'evoluzione della teoria è accompagnata dallo sviluppo del formalismo di descrizione sintattica e tralasciando invece i testi più legati alla psicologia cognitiva e alla filosofia del linguaggio come *Language and Mind* (1972), *Language and Thought* (1993) e il recente *New Horizons in the Study of Language and Mind* (2000). Per una introduzione ai principi e ai metodi della grammatica generativo-trasformazionale si rimanda a RADFORD (1988) (cui ci adeguiamo per la rapida e sommaria esposizione del seguito) e LONGOBARDI (1991), i quali del resto precedono i ripensamenti del *Minimalist Program*, che ha portato Chomsky ad avvicinarsi, forse più implicitamente che esplicitamente, ad altri modelli grammaticali, come la grammatica lessico-funzionale.

Nella grammatica generativo-trasformazionale, la descrizione linguistica a livello sintattico è formulata in termini di analisi dei costituenti. Ciascun costituente è formato da costituenti intermedi oppure da elementi non ulteriormente analizzabili. Un costituente può quindi contenere ricorsivamente un numero imprecisato di livelli intermedi. In questo modo viene a crearsi una gerarchia di costituenti, che può essere rappresentata attraverso una struttura ad albero. I costituenti di livello più basso, cioè le parole, sono suddivise per categoria grammaticale: N (*Noun*), A (*Adjective*), P (*Preposition*), D (*Determiner*), I (*Inflection*), V (*Verb*), Adv (*Adverb*). Alla categoria D appartengono gli articoli e i pronomi dimostrativi, alla categoria I appartengono, ad esempio in inglese e in altre lingue europee, gli ausiliari e i modali. Ciascuna parola è la testa di un costituente³: ad esempio, N è la testa di una NP (*Noun Phrase*), P è la testa di una PP (*Prepositional Phrase*), e così via. Attraverso il meccanismo della ricorsività, un componente XP, che ha per testa la parola di categoria X, può contenere un numero imprecisato di sottocomponenti, tra loro in scala gerarchica. Il modo in cui si costituiscono questi componenti, è dettato da regole di riscrittura del tipo:

³Il fatto che i costituenti siano governati da un sottocostituente che ne costituisce la testa è uno dei temi centrali di CHOMSKY (1981).

5. Analisi sintattica

S(entence) \longrightarrow NP VP

NP \longrightarrow N

...

VP \longrightarrow V NP

...

Ad esempio, *Pers.* 487:

[S[NP[N $\Sigma\pi\epsilon\rho\chi\epsilon\iota\delta\varsigma$]] VP[V $\acute{\alpha}\rho\delta\epsilon\iota$] [NP[N $\pi\epsilon\delta\acute{\iota}\omicron\nu$]]]]

L'esempio è estremamente semplificato e nelle sue regole non esprime, per brevità, le restrizioni dell'accordo di numero, genere e caso. Le regole vanno lette in questo modo: una frase è costituita da un gruppo nominale (il soggetto con eventuali complementi) e da un gruppo verbale (il predicato con eventuali complementi). Il gruppo nominale può essere riscritto come un semplice sostantivo (un'altra regola potrebbe essere: NP \longrightarrow D N, per indicare l'articolo seguito da sostantivo, etc.). Il gruppo verbale può essere riscritto come un verbo seguito da un gruppo nominale, al quale applicare ricorsivamente le regole di NP. Le restrizioni di accordo sono bene espresse in alcune teorie discendenti dalla grammatica generativa, come le *Definite Clause Grammars* (create, fra l'altro, proprio nell'ambito della linguistica computazionale)⁴. La teoria detta *X-bar*, che riguarda i costituenti intermedi, regola e limita il meccanismo della ricorsività. Infatti, se fosse valida la regola NP \longrightarrow D NP, (articolo seguito da gruppo nominale), allora, per ricorsività, sarebbe possibile riscriverla come: NP \longrightarrow D D NP, (articolo seguito da articolo, seguito da gruppo nominale), poi come NP \longrightarrow D D D NP, e così via. Si deve quindi postulare che esistano costituenti intermedi fra N e NP, indicati con N' (più in generale X' fra X ed XP) sui quali agisce il principio della ricorsività rispetto a certe specificazioni, mentre non agisce rispetto a certe altre. Ad esempio, le tre regole seguenti:

NP \longrightarrow D N'

N' \longrightarrow N' PP

N' \longrightarrow N PP

⁴Si veda ad esempio GAZDAR E MELLISH (1989).

lette dal basso verso l'alto, ci dicono che, ad esempio, un nome può essere seguito da un complemento di specificazione, che questo gruppo intermedio può essere seguito da un numero indeterminato di ulteriori specificazioni, ma che il gruppo risultante può essere preceduto da un solo articolo. Regole del tipo $X' \rightarrow X YP$ riguardano i complementi (*complements*) di X , cioè elementi essenziali per X mentre regole del tipo $X' \rightarrow X' YP$ riguardano gli specificatori (*adjuncts*) di X , cioè elementi non essenziali.

La grammatica generativo-trasformativa, almeno nella sua forma originale⁵, distingue due livelli di struttura sintattica: la *d(eep)-structure* e la *s(urface)-structure*. Per passare dal primo livello, concettuale, al secondo, fonologicamente realizzato, intervengono delle regole di trasformazione, che producono, nella *s-structure*, degli spostamenti, delle soppressioni o delle duplicazioni di componenti della *d-structure*. Un esempio può essere la topicalizzazione, la quale porta all'inizio della frase, al fine di metterlo in evidenza, un elemento che nella struttura profonda dovrebbe essere altrove.

Osservazioni sulla grammatica generativo-trasformativa

La grammatica generativo-trasformativa si propone, almeno nelle intenzioni di Chomsky, come grammatica universale; tuttavia, al di fuori di test di laboratorio, la sua applicazione è avvenuta su grandi *corpora* quasi esclusivamente nell'ambito della lingua inglese, come si vedrà meglio nel seguito del capitolo. In inglese, per passare dalla *d-structure* alla *s-structure*, spesso non è necessario l'impiego di un numero elevato di trasformazioni. Invece, in lingue come il greco, dove, a livello

⁵Nel *Minimalist Program* Chomsky rivede le sue teorie precedenti e, anziché affidare l'interpretazione sintattica al dualismo di struttura superficiale e struttura profonda, preferisce concentrarsi su due livelli: A-C (*articulatory-perceptual*) e C-I (*conceptual-intentional*). Al di là della nomenclatura, sembra assumere ora ruolo centrale, nel pensiero del linguista americano, la dicotomia saussuriana fra significante e significato CHOMSKY (1995, p. 169): “[Universal Grammar] must specify the interface levels (A-P, C-I), the elements that constitute these levels, and the computations by which they are constructed. A particularly simple design for language would take the (conceptually necessary) interface levels to be the only levels. That assumption will be part of the 'minimalist' program I would like to explore here” .

5. Analisi sintattica

di *s-structure*, l'*ordo verborum* è più libero e dove è usuale sottintendere elementi essenziali della frase, come il soggetto, la copula, etc., il ricorso alle regole di trasformazione per produrre l'albero sintattico della *d-structure* diventa molto pesante. Tuttavia, durante gli anni Settanta sono comparsi studi per vagliare l'adeguatezza della grammatica generativo-trasformativa per descrivere i fenomeni sintattici della lingua greca. Come prime tappe del dibattito si possono citare PFISTER (1969), contestato da KLOWSKI (1971), seguito dalla replica dello stesso PFISTER (1972). Questi primi articoli si soffermavano su aspetti generali della teoria, come l'applicabilità della nozione di *competence*, caposaldo della dottrina di Chomsky, alle lingue classiche, prive di parlanti nativi. Sono invece l'articolo di KURZOVÁ (1972) e il contributo di EGLI (1973) a entrare nel merito della rappresentazione delle strutture sintattiche del greco per mezzo del formalismo messo a disposizione dalla grammatica generativo-trasformativa. Come fa notare MARTÍNEZ HERNÁNDEZ (1997, pp. 64-65), al quale si rimanda per ulteriori notizie bibliografiche:

SCHMIDT (1979), junto con HOBBS E LAWRENCE (1978), son los autores del único intento realizado hasta ahora de aplicación de una gramática generativo-transformativa completa al griego antiguo.

La grammatica lessico-funzionale

La *lexical functional grammar*⁶ (*LFG*) deriva dalla grammatica generativa, ma rifiuta l'idea di un doppio livello di rappresentazione sintattica, superficiale e profondo, e di conseguenza l'idea che si passi dalla *d-structure* alla *s-structure* (in fase di produzione degli enunciati, o dalla *s-structure* alla *d-structure* in fase di interpretazione) per mezzo di trasformazioni, cioè di movimenti di costituenti. A livello di analisi dei costituenti, vi è quindi un solo schema di rappresentazione, detto *c-structure* (*constituent structure*), corrispondente all'albero sintattico della struttura superficiale chomskyana. Alla base della *LFG* sta la dissociazione fra struttura sintattica e struttura degli argomenti del predicato, in quanto que-

⁶Per una introduzione, si vedano BRESNAN (1982), BRESNAN (2001), DALRYMPLE (1999, pp. 1-37) e, per il francese, ABEILLÉ (1993).

st'ultima è codificata a livello di lessico. La base di dati lessicale registra quindi tante entrate quante sono le strutture argomentali associabili a ciascun termine: è tenuta distinta ad esempio l'accezione transitiva da quella intransitiva di uno stesso verbo. Gli argomenti della predicazione hanno un valore semantico, non sintattico: ad esempio per un verbo transitivo sono *agent* e *theme*. Nella *c-structure* questi argomenti possono trovare corrispondenza nelle funzioni sintattiche di *subject* e *object*, nel caso della forma attiva; tuttavia, nel caso della forma passiva, l'*agent* può anche non essere espresso, avere quindi valore 0, mentre il *theme* va a ricoprire la funzione di *subject*. Così dal punto di vista della interpretazione grammaticale non vi è più una forma originaria (codificata nella struttura profonda) e una struttura derivata per via di trasformazione sintattica (codificata nella struttura superficiale). La forma passiva non è fatta derivare da una più originaria forma attiva, ma si mappa invece, a livello di costituenza sintattica, in maniera diversa rispetto alla forma attiva sulla medesima struttura argomentale, che ha valore semantico ed è insita nel lessico. Del resto, si è visto sopra come Chomsky stesso abbia progressivamente ridotto la distanza, fino a farla quasi scomparire nel *Minimalist Program*, fra struttura profonda, originaria, e struttura superficiale, derivata. Oltre alla *c-structure*, cioè la rappresentazione dell'albero dei costituenti sintattici, la *LFG* prevede una seconda forma di rappresentazione, la *functional structure* (*f-structure*), concepita come una matrice di coppie attributo-valore volta a combinare l'informazione contenuta nel lessico (come la struttura argomentale) con l'informazione prodotta dall'analisi dei costituenti (come le funzioni grammaticali di soggetto e oggetto). La *LFG* si basa quindi sulla codifica delle regole d'impiego della *langue* nella base di dati lessicale e della codifica della struttura formale della *parole* nell'analisi dei costituenti. La *LFG* e le sue derivate, in particolare la *Functional-Unification Grammar*⁷, sono fra i modelli che trovano maggior consenso in linguistica computazionale⁸, proprio per la possibilità di dedurre, per mezzo di algoritmi appropriati e regole di

⁷Si veda ad esempio KAY (1985).

⁸Si vedano ALLEN (1995), NORVIG (1992, pp. 684-750) e DELMONTE (1992).

5. Analisi sintattica

inferenza, la struttura sintattico-semantiche della frase a partire dalla base di dati del patrimonio lessicale.

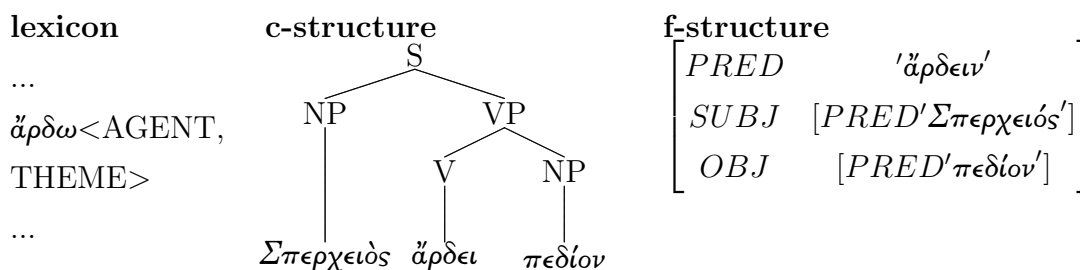


Figura 5.1: Rappresentazione semplificata in *LFG* (*Pers.* 487)

La grammatica di dipendenza

Anche la grammatica di dipendenza, che trae origine dalle idee di TESNIÈRE (1988), propone un modello che non prevede la separazione di una struttura profonda da una struttura di superficie. Presuppone invece l'esistenza di quattro grandi categorie che si connettono fra di loro. Tali categorie sono: **O** (funzione abituale dei sostantivi), **A** (funzione abituale degli aggettivi), **I** (funzione abituale dei verbi) ed **E** (funzione abituale degli avverbi). Queste categorie hanno valore dal punto di vista funzionale e strutturale, non morfologico. Ogni vocabolo, che sul piano paradigmatico appartiene a una determinata categoria, sul piano sintagmatico può ricoprirne una diversa. Così, ad esempio, un aggettivo sostantivato fa parte della categoria **O**, un sostantivo al caso genitivo, dato che ricopre un ruolo di specificazione come gli aggettivi, fa parte della categoria **A**, e così via. Il passaggio da una categoria all'altra viene effettuato per mezzo dei traslativi. Le preposizioni, ad esempio, sono dei traslativi in quanto sono in grado di trasformare, funzionalmente, il sostantivo al quale si accompagnano in un nodo che ricopre un ruolo avverbiale. Il caso genitivo, come si è visto, è un traslativo che trasforma componenti di categoria **O** in componenti di categoria **A**. Una congiunzione subordinante, ad esempio *ἐπεὶ*, trasforma l'intera subordinata in un nodo che, funzionalmente, ricopre il ruolo di un avverbio. Esistono poi i giuntivi, come le congiunzioni coordinanti, che collegano fra di loro nodi appartenenti

alla medesima categoria e allo stesso livello sintattico. Tutte le parole della frase stabiliscono fra di loro dei legami gerarchici, a seconda delle categorie alle quale vengono ad appartenere in seguito alle eventuali traslazioni, e ai livelli stabiliti dalle giunzioni. Gli elementi di tipo A, ad esempio, sono subordinati, in quanto specificano gli elementi di tipo O, gli elementi di tipo O e di tipo E sono subordinati agli elementi di tipo I. Al livello più alto della gerarchia, quindi, sta la funzione del verbo. Il verbo può avere diverse valenze, a seconda del numero di attanti che entrano in gioco: ad esempio, i verbi intransitivi sono monovalenti, in quanto hanno un solo attante, il soggetto; i verbi transitivi sono bivalenti, in quanto hanno due attanti, il soggetto e l'oggetto; i verbi di dare e di dire hanno tre attanti, il soggetto, l'oggetto e il termine. Per traslazione, una proposizione subordinata può ricoprire il ruolo di un attante, come nel caso delle proposizioni oggettive o soggettive. Il verbo delle proposizioni subordinate è connesso superiormente, sull'albero sintattico, con il verbo della reggente e inferiormente con i suoi attanti e circostanti. L'albero sintattico che rappresenta la struttura descritta da una grammatica di dipendenza individua come nodi soltanto elementi presenti nella frase effettivamente realizzata: non vi sono, come negli alberi sintattici della grammatica generativo-trasformazionale, dei nodi che rappresentino livelli intermedi e astratti fra la frase e i suoi costituenti. Tesnière, inoltre, per esemplificare la generalità del suo modello, prende in considerazione molte lingue, anche tipologicamente differenti, dal greco al basco, dal cinese all'inglese, dal turco al russo, etc. La linguistica computazionale in ambito tedesco, ad esempio, preferisce avvalersi del modello della grammatica di dipendenza (o di suoi derivati) piuttosto del modello generativo-trasformazionale, come si può vedere in BRÖKER (1998).

Ai fini della presente ricerca, è da notare che TESNIÈRE (1988, p. 632) indica esplicitamente, fra gli usi del suo modello e del relativo albero sintattico (stemma), quello per lo studio dello stile:

[...] le stemma a l'avantage de permettre de se rendre compte explicitement des caractéristiques de style contenues implicitement dans un passage donné et que les gens ayant le sentiment correct et délicat des finesses de leur langue maternelle sentent instinctivement. Il permet ainsi de préciser la nature exacte des ressources que la structure syntaxique d'une langue

5. Analisi sintattica

donnée met à la disposition des écrivains et des procédés stylistiques dont le judicieux emploi peut enrichir leur art.

Un'applicazione della sintassi di dipendenza a testi latini

Nell'ambito delle lingue classiche, un'applicazione del modello elaborato da Tesnière è stata fatta da ÉVRARD (1989). Partendo da un archivio di testi del C.I.P.L.-L.A.S.L.A., Évrard ha attribuito a ciascuna parola il numero di riferimento della parola da cui è retta. Così, ha marcato gli aggettivi con il riferimento al sostantivo di cui sono attributi, i sostantivi con il riferimento ai verbi da cui sono retti, etc. Il verbo della principale non fa riferimento a nulla (0). In questo modo, all'interno del modello di Tesnière, considerando che una parola è retta da una sola altra parola e può reggerne nessuna, una o più di una, è possibile, attraverso il solo riferimento alla parola reggente, ricostruire l'albero sintattico. Si è detto che una parola può avere un unico reggente, ma esistono delle eccezioni, contemplate da Tesnière stesso e messe in evidenza da Évrard, rappresentate dalle congiunzioni e dai pronomi relativi, che necessitano, secondo il metodo di Évrard, di un raddoppiamento dovuto alla loro doppia funzione. Sembra opportuno riportare le considerazioni di Évrard sull'opportunità di questo raddoppiamento, (p. 116), anche perché costituisce un punto di differenza importante rispetto al modello adottato nella presente ricerca:

Considérons d'abord le cas du dédoublement avec jonction. Il s'agit du cas où plusieurs mots remplissent la même fonction par rapport à un même régissant. Le plus souvent, il sont unis par une conjonction, mais il peuvent aussi être en asyndète. Dans le premier cas, Tesnière estime que les deux mots sont également en connexion avec le régissant, tandis que la conjonction, comme jonctif, a une double connexion avec chacun d'entre eux. Je crois préférable de considérer les faits de la manière suivante: le premier dépendant est subordonné au régissant; il régit la conjonction, laquelle régit le second élément. Soit Cic., *Tusc.*, I, 43: *cum ... naturam sui similem contigit et adgnovit*. L'auditeur, qui structure la phrase mot après mot, reconnaît en *contigit* un verbe régi par *cum*. Le *et* l'avertit qu'il y a une "rallonge" représentée par *adgnovit*, lequel est, en fait, concaténé à l'ensemble par *et*. La situation est différente s'il y a asyndète: il me semble qu'alors, le régi, si cela se trouve, fonctionne deux fois; soit encore Cic.,

Tusc., I, 54: *hic fons, hoc principium est mouendi*. Il est clair que *mouendi* est régi et par *fons* et par *principium*. Il a donc une double fonction; comme il n'y a pas de concaténation par *et*, il conviendra sans doute de le dédoubler. Le dédoublement est particulièrement clair lorsqu'il s'agit du relatif [...] d'une part, en tant que [...] translatif [...] il exerce une fonction translative sous la dépendance de son antécédent, par l'intermédiaire duquel il introduit, avec translation, la relative dans la structure de niveau supérieur; d'autre part, en tant qu'anaphorique, il est régi directement ou indirectement par le verbe de la proposition relative, lequel, en revanche, est concaténé à la structure d'ensemble par le fait qu'il dépend du relatif en tant que translatif.

Si riprenderanno in seguito alcuni punti di queste osservazioni. Évrard, dopo aver marcato il testo in questo modo, raddoppiando le congiunzioni e i pronomi relativi e segnando a fianco di ciascuna parola il riferimento alla parola reggente, fa in modo che il calcolatore segni la distanza, in termini di numero di parole intercorrenti, fra la parola reggente e la parola retta. Tale distanza può essere 0, nel caso del verbo della reggente, che non è retto da nulla, oppure un numero positivo se la reggente segue la parola subordinata o negativo se la reggente precede la parola subordinata. Tramite queste procedure, Évrard è in grado di fornire dati statistici sull'*ordo verborum* preferenziale, sul numero di componenti dei vari sintagmi, etc.

Osservazioni sulla grammatica di dipendenza

La grammatica di dipendenza si rivela un modello di descrizione dei fenomeni sintattici più adeguato a descrivere una lingua con l'ordine delle parole libero come il greco. Tuttavia, una struttura dove ciascuna parola sia in relazione soltanto con una singola parola da cui dipende gerarchicamente (tranne nelle eccezioni viste sopra), impedisce la possibilità di individuare sottostrutture rappresentate dai costituenti. Un costituente, cioè un gruppo di parole ampliabile ricorsivamente per formare sintagmi più articolati, mantiene all'interno della frase una propria identità. Ad esempio, un pronome può fare riferimento proprio a un costituente, piuttosto che a una singola parola. Un costituente può rappresentare una descrizione definita in senso russelliano: si veda ad esempio la descrizione definita $\hat{\omega}$

5. Analisi sintattica

γεραιὰ μῆτερ ἡ Ξέρξου φίλη di *Pers.* 832, che ha per referente la regina Atossa. Sembra più appropriato pensare che sia il costituente *μῆτερ ἡ Ξέρξου*, piuttosto che il solo termine *μῆτερ* a formare l'unità referenziale: prova ne sia che, applicando il test di sostituibilità per determinare la costituenza⁹, una parafrasi come *ὦ γεραιὰ Ἄτσοσα φίλη* avrebbe la medesima denotazione (benché senso alquanto diverso) del precedente *ὦ γεραιὰ μῆτερ ἡ Ξέρξου φίλη*. Al contrario, la sostituzione di *Ἄτσοσα* al solo *μῆτερ* darebbe un inaccettabile *ὦ γεραιὰ Ἄτσοσα ἡ Ξέρξου φίλη*.

5.1 La linguistica dei *corpora*

Lo studio di *corpora* molto estesi di testi scritti o di trascrizioni di messaggi orali gioca un ruolo di primo piano nella linguistica computazionale applicata alle lingue moderne. La linguistica computazionale si affianca in questo modo alla linguistica tradizionale fornendole strumenti adeguati per verificare le teorie grammaticali su grandi quantità di enunciati reali e non su un numero limitato di esempi costruiti *ad hoc* per giustificare la teoria stessa¹⁰. Si viene a creare quindi un rapporto di circolarità virtuosa fra il formalismo applicato durante la fase di annotazione di un *corpus* e la presenza nel *corpus* di fenomeni linguistici di cui il formalismo adottato non è in grado di rendere ragione. Alla fine del processo infatti il formalismo stesso, cioè l'insieme delle regole di codifica dei fenomeni linguistici presi in esame, si arricchisce e si sviluppa in rapporto al *corpus* al quale viene applicato.

In questo modo di procedere ci sono vistosi punti di contatto rispetto alla consolidata tradizione della grammatica descrittiva otto-novecentesca per i *corpora* dei testi classici, tuttavia le differenze sono di un certo rilievo, e si cercherà brevemente di metterne in luce alcune. In primo luogo, la grammatica descrittiva tradizionale può avvalersi di tutta l'espressività della lingua naturale che funge

⁹Si veda LONGOBARDI (1991, pp. 53–73).

¹⁰Esempio evidente di questo vizio è la grammatica di Port Royal la quale, nella tensione fra interpretazione della struttura logica e descrizione della realtà linguistica, è portata a sacrificare la seconda.

da metalinguaggio (per esempio la lingua tedesca del KÜHNER-GERTH, che funge da metalinguaggio rispetto alla lingua greca). Se il metalinguaggio impiegato è una lingua naturale, allora il suo codice è aperto tanto quanto lo permette il registro adottato per effettuare la descrizione dei fenomeni linguistici presi in esame. Al contrario, il metalinguaggio utilizzato per annotare un *corpus* digitale è un linguaggio formale che risponde a una propria semantica e a una propria sintassi, descritte in un metalinguaggio di secondo livello. Il numero dei descrittori e le modalità di impiego sono quindi rigidamente fissate e un loro ampliamento richiede una modifica esplicita del metalinguaggio.

In secondo luogo, il curatore di una grammatica descrittiva non è tenuto a rendere ragione di tutti i fenomeni che insistono su ogni singola unità del corpus esaminato, ma soltanto di quelli ritenuti notevoli. Questo comporta una dinamica fra norma ed eccezione sbilanciata in favore di quest'ultima, con conseguenze rilevanti. Così, lo *Stellenregister* di una qualsiasi grammatica è molto utile per fenomeni di bassa frequenza, ma di scarsa o di nessuna utilità per fenomeni di alta frequenza. Mentre una concordanza indicizza tutto il patrimonio lessicale di un *corpus*, le grammatiche invece si comportano in modo selettivo. All'opposto, lo scopo principale di un *corpus* digitale annotato con descrittori sintattici è proprio l'applicazione sistematica di marcatori unità per unità, in modo esaustivo. Dalla base di dati così trattata è quindi possibile estrarre un indice di tutti i fenomeni e non soltanto di quelli notevoli.

In terzo luogo, la classificazione omogenea dei fenomeni linguistici costituisce il presupposto per poter condurre indagini statistiche sulla base di dati presa in esame. Quantificare le occorrenze di un fenomeno aiuta a descrivere come si articolano la dinamica fra rispetto e infrazione delle regole a livello sintagmatico, fra norma ed eccezione a livello paradigmatico. Inoltre, anche la distribuzione anomala di fenomeni normali, come ad esempio l'impiego nel dramma da parte di un solo personaggio di uno stilema impiegato dagli altri personaggi con frequenza molto differente¹¹, costituisce per la statistica linguistica un tema di grande interesse.

¹¹Si veda ad esempio ALLISON (2003, pp. 19–44) sulla distribuzione di *-οις/-οισι*.

5. Analisi sintattica

Tuttavia, l'indagine statistica su fenomeni sintattici deve essere condotta tenendo costantemente presente il pericolo di circolarità fra interpretazione in fase di marcatura e interpretazione in fase di valutazione dei risultati della ricerca. Come è noto, il rischio varia a seconda della natura stessa dei testi e in base alla estensione dei marcatori impiegati. Testi che hanno come primo obiettivo la funzione referenziale (per le lingue moderne: articoli di giornale, relazioni scientifiche etc. per le lingue antiche: testi in prosa) mirano a evitare l'ambiguità sintattica e semantica, mentre al contrario i testi poetici la impiegano sia come efficace mezzo espressivo, sia a causa di vincoli formali da rispettare, in particolare metrici, che favoriscono la forzatura della struttura sintattica. Anche la scelta dei marcatori da utilizzare per descrivere i fenomeni sintattici deve essere calibrata in base al rapporto fra precisione e arbitrarietà di interpretazione. Un modello sintattico basato su un albero binario dove ciascun nodo si sviluppi sempre in una testa e un elemento aggiunto è più preciso rispetto a un modello che consenta il raggruppamento di più elementi alla pari.

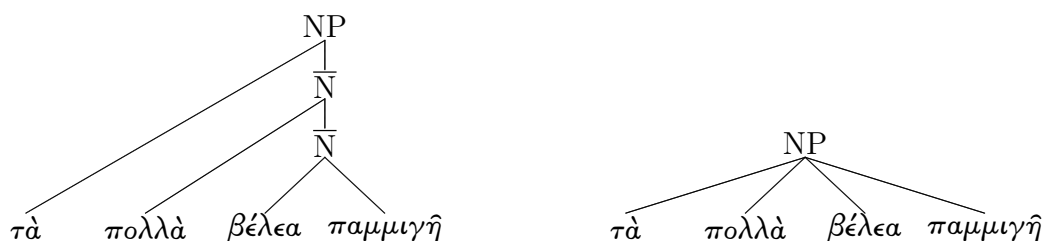


Figura 5.2: Possibili rappresentazioni sintattiche di un sintagma nominale (*Pers.* 269)

Il primo modello però in questo modo è più interpretativo rispetto al secondo e si apre a un maggior grado di arbitrarietà. Lo stesso ragionamento vale per la scelta delle etichette da assegnare agli archi che congiungono i nodi dell'albero sintattico e che descrivono le funzioni da essi ricoperte. È il caso ad esempio del participio congiunto, che può essere marcato in modo più interpretativo, indicandone la natura causale, temporale, etc., oppure meno interpretativo indicandone genericamente la natura circostanziale.

Treebanks

A partire dagli anni Sessanta si sono sviluppati diversi modelli di codifica per la redazione di *corpora* di testi digitalizzati, parallelamente alle metodologie per interrogare le base di dati ed estrarre le informazioni richieste. Il *Brown Corpus*¹², nato in seno all'omonima università americana, è la prima collezione di testi codificati in *readable machine form* con oltre un milione di parole. La collezione iniziale comprendeva testi scritti pubblicati nell'anno 1961, ripartiti per generi di *informative* e *imaginative prose*, mentre nessun genere poetico vi era rappresentato. Nel corso degli anni successivi ciascun *token*¹³ del *corpus* è stato annotato con le informazioni relative alle *parts of speech* (POS) di appartenenza, secondo un sistema di codifica che ha costituito il modello per la redazione di *corpora* successivi¹⁴. Nel corso degli anni Settanta e Ottanta sono state perfezionate le tecniche per eseguire in modo semiautomatico il processo di annotazione delle POS, mentre si elaboravano modelli di codifica della struttura sintattica seguendo le teorie delle *context-free grammars* di derivazione chomskiana.

Con il termine *treebank* si intende un *corpus* annotato con informazioni riguardanti la struttura ad albero dei costituenti sintattici dei periodi contenuti nella base di dati testuale. Uno dei più noti e apprezzati è quello prodotto dal *Penn Treebank Project*¹⁶, sviluppatosi a partire dal 1989 ampliando i materiali del *Brown Corpus* fino a 4.5 milioni di *tokens*, annotati in modo semiautomatico sia per quanto concerne le POS sia per le strutture sintattiche di ciascun periodo. Alla fine della prima fase triennale del progetto i curatori hanno pubblicato in MARCUS ET AL. (1993) i dati sperimentali di un significativo confronto fra la procedura di annotazione interamente manuale delle informazioni morfosintattiche da parte di persone opportunamente addestrate e la procedura semiautomatica, che prevedeva l'attribuzione automatica delle POS e delle strutture sintattiche da parte del calcolatore e una seconda fase di correzione e controllo da parte degli

¹²Si veda, per il sistema di codifica adottato, FRANCIS E KUČERA (1964).

¹³Il termine è usato in linguistica computazionale con un significato simile a quello di *lessia*.

¹⁴Ad esempio l'*International Corpus of English (ICE)*, il *Lundon-Lund Corpus (LLC)* il *Lancaster-Oslo/Bergen Corpus (LOB*¹⁵), etc.

¹⁶<http://www.cis.upenn.edu/treebank>.

5. Analisi sintattica

operatori sui listati prodotti automaticamente. I risultati, in netto favore per la seconda procedura, sono di un certo interesse perché riguardano aspetti sia pratici che teorici del processo di marcatura. Secondo gli autori dell'articolo, le ragioni della superiorità del metodo semiautomatico sono *speed*, *consistency* e *accuracy*. Mentre la velocità di esecuzione, quando non sia modificata sensibilmente di alcuni ordini grandezza, è un fattore che interessa più la politica dei programmi nazionali di ricerca, la consistenza e l'accuratezza hanno delle ripercussioni a livello epistemologico di grande portata (pp. 318–318):

In the first annotation mode (“tagging”), annotators tagged unannotated text entirely by hand; in the second mode (“correcting”), they verified and corrected the output of PARTS¹⁷ [...]. This experiment showed that manual tagging took about twice as long as correcting, with about twice the inter-annotator disagreement rate and an error rate that was about 50% higher. [...]

A simple measure of tagging consistency is inter-annotator disagreement rate, the rate at which annotators disagree with one another over the tagging of lexical tokens, expressed as a percentage of the raw number of such disagreements over the number of words in a given text sample. For a given text and n annotators, there are $\binom{n}{2}$ disagreement ratios (one for each possible pair of annotators). [...]

Consistency, while desirable, tells us nothing about the validity of the annotators' corrections. We therefore compared each annotator's output not only with the output of each other, but also with a benchmark version of the eight texts. This benchmark version was derived from the tagged Brown Corpus by (1) mapping the original Brown Corpus tags onto the Penn Treebank tagset and (2) carefully hand-correcting the revised version in accordance with the tagging conventions in force at the time of the experiment.

Nel seguito dell'articolo gli autori fanno rilevare che risultati del tutto simili si ottengono anche per il processo di annotazione delle strutture sintattiche. Come si vede, i tre parametri di velocità, coerenza e accuratezza aumentano di circa il doppio fra il sistema di annotazione manuale e il sistema di annotazione semiautomatico. Tuttavia, la distinzione fra *consistency* e *accuracy* nasconde un insidioso richiamo al principio di autorità, che è bene prendere in esame. Dal punto di vista logico, la coerenza indica l'assenza di contraddizione fra i risultati dei

¹⁷PARTS è il software per l'assegnazione automatica delle parti del discorso.

diversi operatori. Eppure, nonostante una competenza specifica sia riconosciuta agli operatori, linguisti ai quali è stato impartito un corso di addestramento, tuttavia si ritiene che l'accordo nei risultati del processo di correzione possa essere concordanza in errore indipendente, e per questo si ritiene che l'indice di coerenza non dica nulla sulla validità delle correzioni stesse. L'accuratezza viene misurata ricorrendo all'autorità di una versione di riferimento redatta non dagli operatori ma dagli autori dell'esperimento, ponendo fra i due gruppi una distinzione qualitativa giocata sulla credibilità, dovuta al maggior livello di competenza degli autori rispetto agli operatori e dovuta all'autorevolezza del *Brown Corpus*, dal quale il nuovo *corpus* deriva. Così, mentre *speed* e *consistency* sono parametri oggettivi per misurare la superiorità del procedimento semiautomatico sul procedimento manuale, *accuracy* invece è un parametro legato all'autorevolezza del *benchmark*.

A livello di assegnazione delle POS, le parole possono essere associate a più di una parte del discorso, nel caso in cui vi sia ambiguità intrinseca (per esempio fra alcuni aggettivi e participi della lingua inglese) oppure incertezza da parte dell'operatore. Anche a livello di assegnazione delle strutture sintattiche sono previsti dei meccanismi, di cui si parlerà a breve, per evitare il più possibile le scelte arbitrarie in sede di correzione dei listati prodotti automaticamente.

Penn format

Il *Penn Treebank Project* adotta un sistema di notazione che prevede il raggruppamento dei costituenti sintattici per mezzo di parentesi nidificate, con l'indicazione del tipo di costituente (S, NP, VP, etc.) e della funzione svolta all'interno della frase (-SBJ\verb, -TMP, etc.)¹⁸. Per fare un esempio applicato al greco, seguen-

¹⁸Questo tipo di notazione, detto, a causa dell'abbonante uso di parentesi, *bracketed format*, è largamente impiegato per la redazioni di *treebanks*: si veda ABEILLÉ (2003) per una panoramica di altri progetti che ne fanno uso. Una delle ragioni della sua diffusione è il fatto che questa notazione è adottata nel linguaggio di programmazione LISP per rappresentare la sua struttura dati preferenziale, la *linked list*, che è appunto isomorfa rispetto a una struttura ad albero. Il LISP e il Prolog, proprio per la loro naturale capacità nel gestire questo tipo di strutture, sono stati fino alla fine degli anni Novanta fra i linguaggi prediletti per il trattamento

5. Analisi sintattica

do le linee guida del *Penn format II* esposte in BIES ET AL. (1995), il sintagma *Pers.* 580-1 $\pi\epsilon\nu\theta\epsilon\acute{\iota}$ (δ') $\acute{\alpha}\nu\delta\rho\alpha$ $\delta\acute{o}\mu\omicron\varsigma$ $\sigma\tau\epsilon\rho\eta\theta\epsilon\acute{\iota}\varsigma$ verrebbe rappresentato nel modo seguente¹⁹:

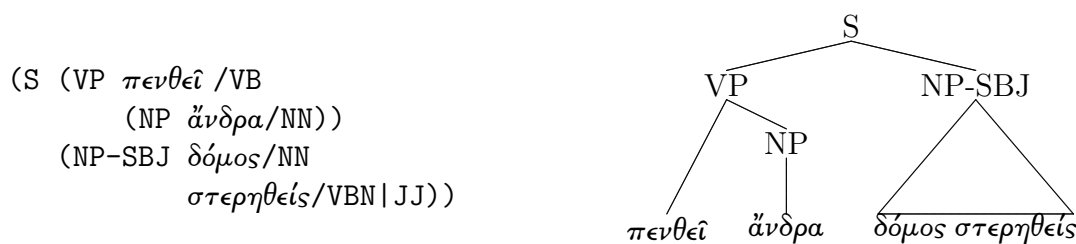


Figura 5.3: *Pers.* 580

Come si vede, la notazione in *bracketed format* riportata sulla sinistra contiene tutte le informazioni necessarie per rappresentare l'albero sintattico raffigurato a destra. Accanto a ciascuna parola è indicata la relativa POS²⁰: VB (*Verb, base form*), NN (*Noun, singular or mass*), VBN (*Verb, past participle*), JJ (*Adjective*). La prima etichetta dopo ciascuna parentesi indica sempre la categoria del costituente: S (*Sentence*), VP (*Verbal Phrase*), NP (*Nominal Phrase*), etc. Separata da una lineetta può essere indicata la funzione del costituente rispetto al costituente che lo domina gerarchicamente. In questo caso SBJ (*Subject*) indica che il costituente $\delta\acute{o}\mu\omicron\varsigma$ $\sigma\tau\epsilon\rho\eta\theta\epsilon\acute{\iota}\varsigma$ è il soggetto del periodo analizzato. Il complemento oggetto, privo dell'etichetta di funzione, è rappresentato come un componente del sintagma verbale, la cui testa è rappresentata dal verbo stesso. Le *guidelines* del *Penn Treebank Project* consentono un moderato margine di scelta nel del linguaggio naturale e nel campo dell'Intelligenza Artificiale. Per indicazioni bibliografiche ed esempi, si vedano NORVIG (1992) per il LISP e STERLING E SHAPIRO (1994) per il Prolog, nonché DELMONTE (1992).

¹⁹Si omette in questo esempio la rappresentazione della particella, perché implica il ricorso a meccanismi di cui si parlerà nei paragrafi seguenti.

²⁰I marcatori delle parti del discorso proposti dal *Penn Treebank Project* in realtà non si limitano a indicare la categoria grammaticale di appartenenza, ma sono dei descrittori completi dei tratti morfologici della lingua inglese, abbastanza inadeguati e insufficienti per altre lingue, con tratti morfologici più complessi, come il tedesco o il greco. Per esempio, i marcatori relativi alla categoria grammaticale del verbo sono soltanto cinque, cioè tanti quante sono le forme del paradigma del verbo inglese.

livello di analisi dei componenti, purché lo si faccia in modo coerente per l'intero *corpus*. Tuttavia, per quanto riguarda il sintagma nominale NP, ad esempio, è preferito l'appiattimento sullo stesso livello dell'articolo, del sostantivo e dell'attributo, piuttosto della distinzione fra l'articolo e un sottocomponente formato dal sostantivo e dall'attributo, secondo le regole della teoria X-barra vista sopra.

Per esplicita ammissione dei curatori del progetto (MARCUS ET AL. (1993, p. 316)), “a major concern of the Treebank is to avoid requiring annotators to make arbitrary decisions”, e i dati sperimentali hanno dimostrato che la distinzione forzata fra *arguments* e *adjuncts* è una delle maggiori cause di rallentamento nel processo di annotazione/correzione da parte degli operatori e soprattutto uno dei principali motivi di incoerenza (p. 323):

Allowing annotators to ignore this distinction when it is unclear (attaching constituents high) increases productivity [...] Informal examination of later annotation showed that forced distinctions cannot be made consistently. As a result of this experiment, the originally proposed skeletal representation was adopted, without a forced distinction between arguments and adjuncts.

I meccanismi offerti per trattare i casi di ambiguità sintattica si articolano su diversi piani, e in particolare:

1) A livello di marcatura delle POS, è possibile annotare più di una categoria grammaticale associata a una singola parola. Questo dipende dalla doppia natura dei marcatori delle POS, che devono rendere conto sia degli aspetti morfologici, come numero, tempo, modo, etc., sia degli aspetti funzionali. In questa maniera si possono separare con una barra verticale diverse categorie grammaticali per indicare, come nell'esempio visto sopra, che un participio (VBN) è impiegato in funzione attributiva (JJ). Questo artificio può essere impiegato quindi nei casi in cui sia in atto il fenomeno della traslazione, cioè quel meccanismo studiato approfonditamente da Tesnière, tanto da costituire un tema portante della sua grammatica strutturale, secondo cui un lessema appartenente a una parte del discorso viene impiegato con la funzione di un'altra parte del discorso: l'aggettivo sostantivato è, come si è visto sopra al paragrafo 5, un esempio di questo meccanismo in atto.

5. Analisi sintattica

2) Per quanto concerne l'analisi dei costituenti, è previsto un marcatore speciale per i costituenti di categoria sconosciuta o incerta. Questo tipo di indicazione è utile soprattutto durante il processo automatico di assegnazione dei *tags* e durante la fase di controllo da parte di un gruppo di operatori che lavorano in modo indipendente. Tuttavia, durante la fase di revisione finale, quando i responsabili del progetto hanno la possibilità di intervenire congiuntamente sull'intero *corpus* per verificarne la coerenza, il numero di componenti marcati in questo modo deve essere ridotto al minimo, se non proprio scomparire del tutto.

3) Anche il punto di connessione di un componente all'albero sintattico è esposto ad ambiguità. Per rendere conto della equivocità delle interpretazioni è stato quindi predisposto il meccanismo dello *pseudo-attachment*, che permette di annotare un riferimento a un componente già connesso all'albero sintattico, connettendolo in un punto diverso del medesimo albero. Mentre le ambiguità trattate nei due punti precedenti possono essere ridotte in fase di revisione dei materiali annotati, soprattutto in base al contesto, l'artificio dello *pseudo-attachment* ha lo scopo di rendere ragione di *permanent predictable ambiguities*²¹, che si verificano quando più di un costituente di livello gerarchico superiore è candidato alla reggenza di un solo costituente subordinato, ma il contesto non permette di sciogliere l'ambiguità.

Vi sono anche tre casi in cui si fa ricorso allo *pseudo-attachment* in contesti non ambigui, e cioè: quando vi siano costituenti condivisi ἀπὸ κοινοῦ, quando le parti di un costituente siano discontinue rispetto all'ordine della frase e quando una intera frase sia dislocata rispetto all'ordine atteso. Poiché questo meccanismo prevede la connessione all'albero sintattico non di costituenti realizzati nella struttura di superficie della frase ma di riferimenti, cioè di puntatori, ai costituenti effettivi, lo *pseudo-attachment* è rappresentato da un indice al costituente effettivo e da un elemento nullo. Gli elementi nulli sono privi di contenuto lessicale, in quanto non si manifestano nella struttura di superficie, tuttavia differiscono tra di loro per il fatto di appartenere a tipologie diverse, in base alla loro funzione. Oltre agli elementi nulli impiegati nello *pseudo-attachment* ve ne sono infatti altri

²¹Si veda MARCUS ET AL. (1993, p. 326).

per indicare l'ellissi oppure le tracce di un movimento di trasformazione come ad esempio, nella lingua inglese, il movimento di costituenti richiesto dalle forme *what, who*, etc. nelle interrogative²².

NeGra format

Il sistema di notazione proposto dal *Penn Treebank Project* si adegua quindi alle regole delle *context-free grammars* di matrice chomskyana, integrandole con degli artifici per rendere ragione dei fenomeni di dipendenza in caso di discontinuità dei costituenti, di ellissi, etc. Il grafo espresso dalla notazione in *bracketed format*, nonostante questi accorgimenti, continua a essere un albero ordinato che non consente accavallamenti fra i rami. Tuttavia, come si è detto più volte sopra, questo modello non è adatto a lingue con ordine delle parole più libero rispetto alla lingua inglese, come fanno rilevare i responsabili del tedesco *NeGra*²³ *Project*. SKUT ET AL. (1997, p. 89):

Due to the frequency of discontinuous constituents in non-configurational languages, the filler-trace mechanism would be used very often, yielding syntactic trees fairly different from the underlying predicate-argument structures.

Così, per la redazione di un *treebank* di lingua tedesca paragonabile per ordine di grandezza al *Penn Treebank* è stato sviluppato un modello che si basa su grafi che consentono accavallamenti fra gli archi (*crossing edges*) e, in casi relativamente poco frequenti, archi secondari (*secondary edges*). Poiché i *secondary edges* connettono i nodi creando dei cicli, le strutture che li contengono contravvengono alla definizione stessa di albero, ma non a quella più generale di grafo.

Le ragioni sono sia di ordine pratico che di ordine teorico. Dal punto di vista pratico, la gestione di un numero elevato di tracce e altri elementi vuoti rallenta la fase di annotazione ed è esposta al pericolo di incoerenza fra gli operatori. Dal

²²Per una panoramica dei movimenti di trasformazione si rimanda a RADFORD (1988, pp. 401–593).

²³NeGra è l'acronimo di *Nebenläufige grammatische Verarbeitung*.

5. Analisi sintattica

punto di vista teorico “a uniform representation of local and non-local dependencies makes the structure more transparent”²⁴. Secondo i curatori la redazione di qualsiasi *treebank* deve rispondere ai seguenti requisiti:

- 1) i fenomeni grammaticali devono essere descritti, non spiegati o interpretati;
- 2) le annotazioni devono essere influenzate il meno possibile da considerazioni specifiche di una particolare teoria;
- 3) lo schema di codifica deve offrire dei mezzi di rappresentazione di tutti i fenomeni che occorrono nel testo, senza rigidità aprioristiche (questo principio è detto *data-drivenness* proprio perché il modello deve adeguarsi alla natura dei dati, evitando che la descrizione dei dati sia forzata per adeguarsi al modello).
- 4) si deve tendere a una separazione chiara fra i diversi livelli di annotazione (esposti nei tre punti seguenti);
- 5) le categorie sintattiche vengono codificate a livello dei nodi del grafo sintattico;
- 6) le funzioni grammaticali vengono annotate come etichette degli archi del grafo sintattico;
- 7) le POS vengono associate a livello delle relative parole.

L’impiego di modelli che prevedano la codifica di tracce e altri elementi nulli è, come si è visto, un ulteriore requisito dei più autorevoli *treebanks* in lingua inglese, in quanto è adeguato a descriverne i fenomeni sintattici. Tuttavia, quando tali modelli siano applicati a lingue con ordine libero delle parole, tale impiego rischia di violare in modo pesante il primo, il secondo e il terzo dei requisiti. L’individuazione infatti della presenza effettiva e della natura degli elementi nulli che, per definizione, non sono manifesti nella struttura di superficie, è frutto di interpretazioni legate alle differenti teorie grammaticali che ne fanno uso. Si deve far notare comunque che una struttura contenente *crossing* e *secondary edges* può essere convertita, mappata su una struttura contenente elementi nulli, e questo permette infatti la conversione dal *NeGra format* al *Penn format* e viceversa. Tuttavia, la conversione da un formato all’altro è un procedimento automatico che non coinvolge le scelte degli operatori addetti alla annotazione manuale: que-

²⁴SKUT ET AL. (1997, p. 89).

sto comporta che gli elementi nulli risultanti dal processo di conversione da un formato all'altro siano impiegati coerentemente e in modo del tutto convenzionale.

In conseguenza di quanto detto sopra, il modello proposto dal *NeGra Project* prevede che ci sia una corrispondenza biunivoca fra parole del periodo da analizzare e nodi terminali (foglie) dell'albero sintattico generalizzato²⁵. Il sistema di annotazione adottato si attiene al principio del rasoio di Ockham per evitare il moltiplicarsi di nodi non necessari: gli elementi nulli sono evitati attraverso l'impiego dei *crossing edges*, come nel caso dell'iperbato, oppure dei *secondary edges*, usati per descrivere i casi di ellissi o di ἀπὸ κοινοῦ. Questi archi secondari collegano così due nodi già connessi ma evitano di postulare una entità fittizia che funga da segnaposto. In caso di incertezza, gli operatori addetti al processo di annotazione sono tenuti a connettere i costituenti in dubbio con i nodi gerarchicamente superiori dai quali dipendano in modo certo, senza creare arbitrari nodi intermedi²⁶. In generale il risultato del processo di annotazione, manuale o semiautomatico, deve essere privo il più possibile di informazioni ridondanti: si deve evitare di annotare ogni tipo di informazione che possa essere dedotta in modo automatico a partire dalle informazioni già possedute. Ad esempio, il nodo che rappresenta una *Prepositional Phrase* raggruppa allo stesso livello sia la preposizione, testa del sintagma, sia le componenti nominali che abitualmente sarebbero raggruppate in un costituente intermedio NP, retto dal nodo PP: si può passare dall'una all'altra struttura deducendo l'informazione dalle POS e applicando una regola d'inferenza che raggruppi automaticamente i componenti della categoria dell'aggettivo JJ e i componenti della categoria del sostantivo NN in un sottocostituente NP, congiungendo invece in modo diretto a PP il componente di categoria IN (cioè la preposizione).

Questa architettura comporta quindi una notevole riduzione delle scelte che devono essere prese dagli operatori in fase di annotazione e correzione e riduce di

²⁵È detto *generalized tree*, ad esempio in ZINSMEISTER ET AL. (2001, p. 3), in quanto è un grafo con meno restrizioni rispetto agli alberi sintattici previsti dalle *context-free grammars*.

²⁶Le grammatiche che seguono la teoria X-barra danno prova di incertezza, e quindi propongono modelli possibilistici, proprio sulla determinazione di nodi intermedi. Si veda LONGOBARDI (1991, p. 75).

5. Analisi sintattica



Figura 5.4: *Pers.* 202

conseguenza la possibilità di incoerenze. Per quanto riguarda l'assegnazione delle etichette delle funzioni grammaticali, anche in questo caso è possibile utilizzare un numero ridotto di marcatori, in quanto distinzioni di livello superiore possono essere dedotte dall'applicazione di regole d'inferenza rispetto alle informazioni di livello inferiore. All'interno di una *Nominal Phrase*, ad esempio, tanto all'articolo, quanto all'aggettivo e al sostantivo connessi direttamente al nodo viene assegnata l'unica etichetta NK (*Nominal Kernel*), per indicare che sono tutti elementi costitutivi del nucleo della *Nominal Phrase*.

I dati del *treebank* sono contenuti all'interno di tabelle di un database relazionale, ma possono essere esportati sotto forma di una singola tabella di solo testo, le cui informazioni sono sufficienti e necessarie a ricostruire il grafo sintattico di partenza. I campi di tale tabella sono i seguenti:

campi obbligatori

- 1) parola o numero di identificazione del nodo
- 2) etichetta della POS associata alla parola o della categoria grammaticale del nodo
- 3) sequenza dei tratti morfologici
- 4) etichetta della funzione grammaticale
- 5) riferimento al nodo reggente immediatamente superiore

campi opzionali

- 6) etichetta della funzione grammaticale associata all'arco secondario
- 7) riferimento al nodo connesso tramite l'arco secondario

La Tab. 5.1 mostra un esempio di codifica in *NeGra exporting format* per rappresentare i dati relativi al grafo sintattico di Fig. 5.5.

	parola/ n. del nodo	POS/ cat. del nodo	tratti morf.	funz. gramm.	nodo regg.
1.	ἄλλους	prn	A.m.pl	OA	500
2.	δ'	ptc	–	DM	500
3.	ὁ	art	N.m.sg	NK	501
4.	μέγας	adj	N.m.sg	CJ	502
5.	καὶ	ccd	–	CD	502
6.	πολυθρέμμων	adj	N.m.sg	CJ	502
7.	Νεῖλος	sub	N.m.sg	NK	501
8.	ἔπεμψεν	vrb	aor.ind.act.3p.sg	HD	500
9.	#502	CAP	–	NK	501
10.	#501	NP	–	SB	500
11.	#500	S	–	–	–

Tabella 5.1: Rappresentazione di *Pers. 33-34* in *Negra export format*

Confronto con il modello di Évrard

Come si vede, la struttura di Tab. 5.1 ricorda, a una prima lettura, la proposta di rappresentazione sintattica di Évrard vista sopra, mentre il grafo di cui è espressione (Fig. 5.5) ha forti analogie con i grafi sintattici generati seguendo le linee guida del *Penn Treebank Project*. Se in Évrard l'aggettivo dipende direttamente dal sostantivo cui fa riferimento, e questi a sua volta dipende direttamente da un altro sostantivo, da un verbo, etc., qui invece i nodi terminali, le parole, dipendono sempre da nodi non terminali, i sintagmi (i quali del resto possono essere rappresentati nella medesima tabella). In effetti, il modello ideato dal gruppo di studio tedesco nasce proprio dalla felice fusione di alcuni principi della grammatica di dipendenza da un lato con l'analisi dei costituenti dall'altro. In BRANTS ET AL. (2003, p. 75) gli autori ammettono che la grammatica di dipendenza era il loro modello iniziale, il quale tuttavia non era soddisfacente per descrivere alcuni fenomeni linguistici di rilievo:

We used Dependency Grammar as the starting point for the development of our annotation scheme. It soon turned out that the sharp distinction

5. Analisi sintattica

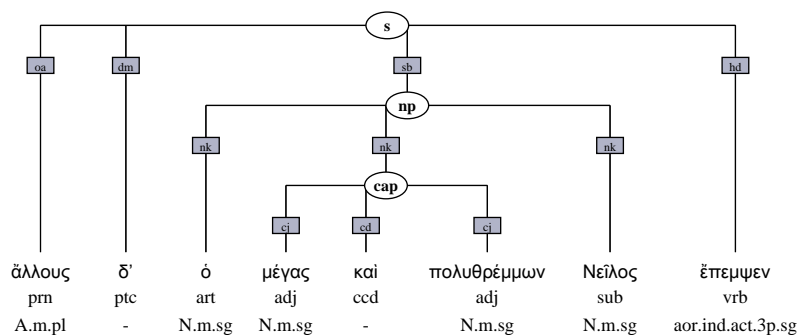


Figura 5.5: Grafo sintattico relativo ai dati di Tab. 5.1

between heads and modifiers stipulated by the formalism causes difficulties in practice. In particular, all kinds of constructions without a clear syntactic head are difficult to analyse: ellipses, sentences without a verb (e.g., copula-less predicatives), and coordinations.

Se la grammatica di dipendenza offre uno schema di annotazione valido e ben sperimentato per le lingue con ordine libero delle parole²⁷, tuttavia la descrizione di alcuni fenomeni linguistici richiede delle forzature. Si è visto sopra come Évrard debba duplicare nelle sue tabelle i pronomi relativi e le congiunzioni per indicarne la duplice funzione, mentre il sistema di annotazione del *Prague Treebank* prevede, ad esempio, che due termini coordinati per asindeto abbiano come testa sintattica addirittura il segno di interpunzione che li separa²⁸.

²⁷Sistemi di annotazione basati sulla grammatica di dipendenza, oltre che per il latino, sono stati adottati sia per le lingue romanze sia per le lingue slave. Si veda ABEILLÉ (2003), nelle sezioni corrispondenti.

²⁸Questo nella critica mossa da BRANTS ET AL. (2003, p. 76); si vedano però esempi di annotazione della coordinazione e dell'ellissi secondo il *Prague Treebank* in BÖHMOVÁ ET AL. (2003, pp. 110–112).

5.2 Dall'interesse linguistico all'interesse filologico

Riprendendo quanto detto all'inizio di questo capitolo, i modelli esistenti per la redazione di *treebanks* non sono finalizzati alla creazione di repertori di interpretazioni sintattiche storicamente determinate. Questo comporta la mancanza di costrutti specifici per trattare lezioni alternative rispetto ai sintagmi analizzati, interpretazioni alternative della struttura e della funzione del medesimo sintagma, riferimenti bibliografici relativi alla lezione e/o all'interpretazione sintattica. Di seguito si riporta un esempio per mostrare come le varianti o le congetture meriterebbero un trattamento specifico e come sul medesimo testo possano insistere più interpretazioni sintattiche.

Murray, West e la maggior parte degli editori moderni accolgono nel testo di *Pers.* 168 la congettura di Heimsoeth, che sostituisce, sulla base dell'interpretazione di uno scolio, *ὀφθαλμοῖς* della paradosi con *ὀφθαλμῶ*²⁹: *ἔστι γὰρ πλοῦτός γ' ἀμεμφής, ἀμφὶ δ' ὀφθαλμῶ φόβος*.

L'apparato di West dà la notizia bibliografica completa della congettura accolta nel testo: Heimsoeth³ 72, cioè, come si ricava dall'*Index criticorum* della sua edizione, F. Heimsoeth, *Kritische Studien zu den griechischen Tragikern*, Bonn 1865, p. 72. Per questa lezione, il valore di *ἀμφὶ ... ὀφθαλμῶ* inteso come complemento di argomento è discusso in MATINO (1998, p. 71) e tale interpretazione sintattica è espressa nel grafo di Fig. 5.6.

Invece, per la lezione dei codici *ὀφθαλμοῖς* (tranne Q, che ha un aberrante *ὀφθαλμός*), la Matino registra sia l'interpretazione sintattica proposta da Lalin (stato in luogo figurato), sia quella proposta dalla Bile (causale-finale). Le due interpretazioni sintattiche sono rappresentate rispettivamente nel primo e nel secondo riquadro di Fig. 5.7.

²⁹Si tralascia qui di discutere le profonde implicazione sul piano del referente: occhi della regina, oppure metafora che indica Serse, occhio della casa.

5. Analisi sintattica

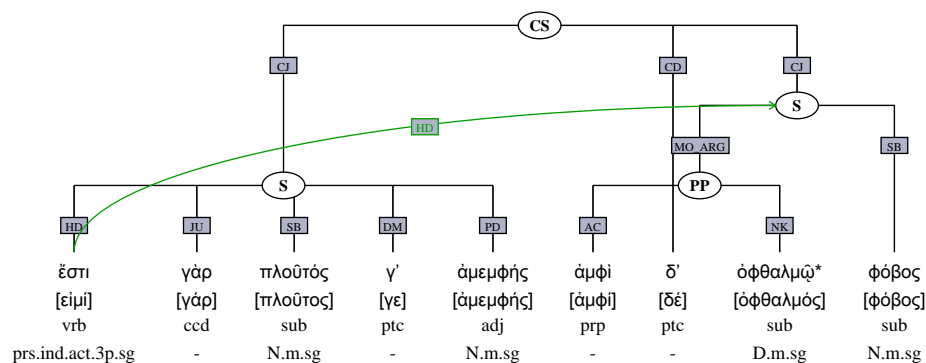


Figura 5.6: *Pers.* 168 (editori moderni)

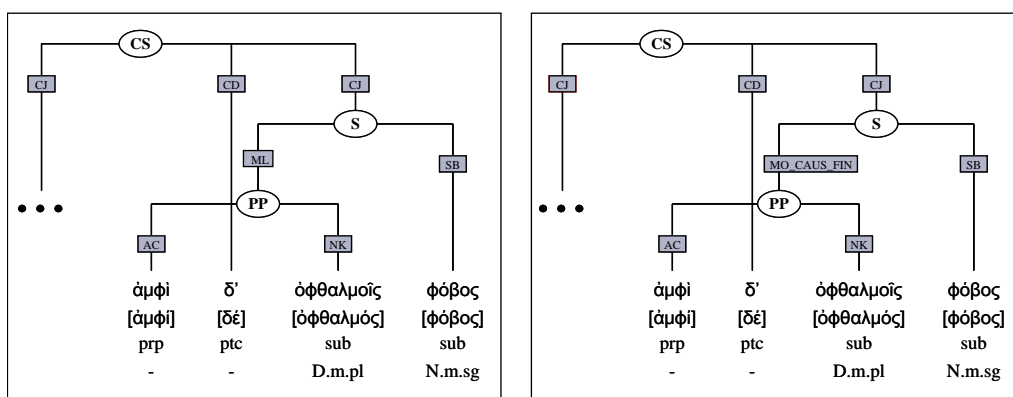


Figura 5.7: Due interpretazioni sintattiche di *Pers.* 168 (paradosi)

5.3 Ancient Greek Dependency Treebank

Il *Perseus Project* ha adottato due diversi modelli protocolli di annotazione per la creazione dell'*Ancient Greek Dependency Treebank*³⁰, descritti in BAMMAN ET AL. (2009) e in MAMBRINI (2011). Secondo il protocollo standard, con cui sono stati annotati i poemi omerici, il testo viene analizzato da due annotatori indipendenti e, in caso di disaccordo, viene armonizzato da un supervisore. Se-

³⁰http://perseusdl.github.io/treebank_data

5.3 Ancient Greek Dependency Treebank

condo il modello autoriale invece, adottato per le tragedie di Eschilo e di Sofocle, il testo viene annotato da un singolo specialista.

I testi annotati sono stati usati per studi linguistici (MAMBRINI E PASSAROTTI, 2013) e linguistico-stilistici (MAMBRINI E PASSAROTTI, 2016). Su parte del testo di Tucidide MAMBRINI (2013) ha dimostrato l'utilità dell'annotazione secondo tutti i *layers* previsti dalla *Dependency Grammar*: testuale, morfologico, analitico (sintattico) e tectogrammaticale (annotazione dei ruoli semantici).

L'*Open Philology Project* coordina il lavoro di annotazione di testi classici grazie alla cooperazione di studenti. Allo scopo sono state redatte delle linee guida (CELANO, 2014).

Per garantire l'interoperabilità fra le risorse testuali annotate sintatticamente, si sta convergendo verso *Universal Dependencies*³¹ (NIVRE ET AL., 2016).

³¹<http://universaldependencies.org>

6

Modelli e risorse lessico-semantici

Gli obiettivi dell'analisi semantica nell'ambito della presente ricerca sono molteplici: in primo luogo ci si propone di studiare la distribuzione del lessico in rapporto al piano del contenuto e non solo al piano dell'espressione, prendendo in considerazione la polisemia e la sinonimia, nonché i rapporti attivi sul piano paradigmatico dell'iperonimia e dell'iponimia, dell'olonimia e della meronimia. In secondo luogo, sul piano sintagmatico, interessa studiare la cooccorrenza di tratti semantici per valutare l'articolazione delle isotopie presenti nel testo. In terzo luogo si cerca di valutare il rapporto fra sintassi e semantica, per esaminare coerenze e incoerenze fra struttura e contenuto espresso.

In generale, si può condividere la critica di MARTÍNEZ HERNÁNDEZ (1997, p. 25)¹, che valuta il rapporto fra studi classici e moderne teorie linguistiche nell'ambito della fonologia, della sintassi e della semantica concludendo:

[...] donde el retraso resulta impresionante es en el dominio de la semántica. A este respecto constatamos que la ciencia que se ocupa del lado del contenido de la lengua, el más importante, a nuestro entender, es la que más marginada ha quedado en filología griega, en comparación con las disciplinas que se ocupan del lado formal. Así como en las distintas lenguas

¹Oltre a questo prezioso sguardo d'insieme sugli studi di semantica del greco antico, è necessario ricordare, a cura dello stesso studioso (MARTÍNEZ HERNÁNDEZ, 2000), la pubblicazione degli atti del Congresso internazionale di semantica, tenutosi presso l'Università di Laguna nel 1997 in occasione del centenario dell'*Essai de sémantique* di M. Bréal. Insieme a studi di semantica delle lingue moderne, una sezione è interamente dedicata al greco.

6. Modelli e risorse lessico-semantici

modernas disponemos de varios manuales de semántica para cada una, para el griego antiguo no tenemos todavía una obra comparable en la que se pueda recoger y sistematizar los logros parciales conseguidos ya en los diversos trabajos prácticos. [...] Hay quien ha querido ver la causa de este atraso en semántica griega en el hecho de que se la ha considerado como una disciplina híbrida a caballo entre la filología y la lingüística, por lo que es ignorada o subestimada por los representantes de estos dos dominios.

È sufficiente consultare la ricca bibliografia proposta da Martínez Hernández stesso, integrata dalla rassegna di PERDICOYIANNI-PALÉOLOGUE (2003), per avere una conferma sulla frammentarietà degli studi di semantica greca, orientati per lo più a indagare nel dettaglio singoli campi semantici, contenenti numeri molto ristretti di membri.

6.1 La semantica strutturale

Considerazioni preliminari sulla prospettiva strutturalista in semantica

Alla fine del '67, un anno dopo la pubblicazione di *Sémantique structurale* di Greimas, Italo Calvino tiene a Torino e in altre città d'Italia e d'Europa una conferenza dal titolo suggestivo di *Cibernetica e fantasmi*, in cui lo scrittore mette in luce la convergenza fra scienze umane e scienze esatte nell'impostazione metodologica, e si potrebbe dire ideologica, di quegli anni²:

Nel modo in cui la cultura d'oggi vede il mondo, c'è una tendenza che affiora contemporaneamente da varie parti: il mondo nei suoi vari aspetti viene visto sempre più come *discreto* e non come *continuo*. Impiego il termine "discreto" nel senso che ha in matematica: quantità "discreta" cioè che si compone di parti separate. Il pensiero, che fino a ieri ci appariva come qualcosa di fluido, evocava in noi immagini lineari come un fiume che scorre o un filo che si sdipana, oppure immagini gassose, come una specie di nuvola, tant'è vero che veniva spesso chiamato "lo spirito", – oggi tendiamo a vederlo come una serie di stati discontinui [...]

²I passi seguenti sono tratti da CALVINO (1995, pp. 203-205).

L'origine di questo mutamento è visto negli studi sulla teoria dell'informazione³ e sulla cibernetica dei due decenni precedenti:

Shannon, Weiner, Von Neumann, Turing, hanno cambiato radicalmente l'immagine dei nostri processi mentali. Al posto di quella nuvola cangiante che portavamo nella testa fino a ieri e del cui addensarsi o disperdersi cercavamo di renderci conto descrivendo impalpabili stati psicologici, umbratili paesaggi dell'anima, – al posto di tutto questo oggi sentiamo il velocissimo passaggio di segnali sugli intricati circuiti che collegano i relé, i diodi, i transistor di cui la nostra calotta cranica è stipata.

Tuttavia Calvino invita a vedere questo processo che porta dal continuo al discontinuo non come una riduzione, un impoverimento del primo rispetto al secondo. Il discontinuo, anzi, permette di organizzare e descrivere la complessità del continuo.

Potrei dire anche che la numerabilità, la finitudine, stanno avendo la meglio sull'indeterminatezza dei concetti che non possono essere sottoposti a misurazione e a delimitazione, ma questa formulazione rischia di dare un'idea un po' semplicistica di come stanno le cose, mentre è vero proprio tutto il contrario: ogni processo analitico, ogni divisione in parti tende a dare del mondo un'immagine che si va via via complicando, così come Zenone da Elea rifiutandosi d'accettare lo spazio come continuo finiva per aprire tra la tartaruga e Achille una suddivisione infinita di punti intermedi. Ma la complicazione matematica può essere digerita istantaneamente dai cervelli elettronici.

Nel passo successivo l'autore sembra propendere per un modello dove la finitezza e la discontinuità non si applicano soltanto a livello interpretativo ma sono insite nella struttura stessa della realtà:

Il processo in atto oggi è quello d'una rivincita della discontinuità, divisibilità, combinatorietà, su tutto ciò che è corso continuo, gamma di sfumature che stingono una sull'altra. Il secolo decimonono, da Hegel a Darwin, aveva visto il trionfo della continuità storica e della continuità biologica che superava tutte le rotture delle antitesi dialettiche e delle mutazioni genetiche. Oggi questa prospettiva è radicalmente cambiata [...]

³Tuttavia RASTIER (2001, *passim*) fa notare che la teoria di Shannon, nata per studiare il modo in cui ottimizzare la trasmissione a distanza di segnali, si occupa del piano dell'espressione e non del piano del contenuto.

6. Modelli e risorse lessico-semantiche

Così, un numero limitato di opposizioni o di regole di costruzione è sufficiente a rendere ragione della complessità dei fenomeni, tanto fisici quanto mentali. L'autore vede orientate verso questa prospettiva anche le scienze del linguaggio, non soltanto nella rappresentazione del significante, secondo un processo aperto da de Saussure già nei primi decenni del secolo e portato a piena maturazione da Jakobson, ma anche nella rappresentazione del significato:

[...] la linguistica strutturale tende a configurarsi in un gioco d'opposizioni altrettanto semplice che la teoria dell'informazione [...]

Un'estrema semplificazione di formule logiche è usata anche dalla scuola francese della semantica strutturale di A.J. Greimas [...]

Questo quadro, che ci è sembrato utile riportare per esteso in quanto fornisce alcune coordinate culturali del clima in cui si è formato *Sémantique structurale*, era ben presente già allo stesso Greimas⁴:

La linguistique traditionnelle – qui se conformait d'ailleurs en ceci aux tendances générales de son temps – insistait volontiers sur le caractère continu des phénomènes linguistiques. Ainsi, le passage de *a* latin de *mar* à [ε] français de *mer* était considéré comme inconscient, insaisissable, sans solution de continuité [...]

C'est dans ce contexte qu'il faut situer, en lui rendant son caractère révolutionnaire, l'affirmation saussurienne que la langue est faite d'oppositions. [...]

La seule façon d'aborder, à l'heure actuelle, le problème de la signification consiste à affirmer l'existence de discontinuités, sur le plan de la perception, et celle d'écart différentiels (ainsi Lévi-Strauss), créateurs de signification, sans se préoccuper de la nature des différences perçues.

Debitore dell'analisi componenziale⁵ post-saussuriana, in particolare di Hjelmslev, Jakobson e, più nello specifico, di Pottier⁶, il sistema di Greimas si fonda su opposizioni binarie.

⁴GREIMAS (1995, p. 18).

⁵Si veda LYONS (1977, pp. 317–335), dove è tracciata una panoramica sull'argomento, con riferimenti bibliografici.

⁶Non si devono dimenticare tuttavia gli influssi che ha avuto, prima e dopo *Sémantique structurale*, la teoria del *Wortfeld*, per la quale si può vedere TRIER (1973), nello sviluppo della semantica strutturale, in particolare nella speculazione di Coseriu, come più volte ricorda MARTÍNEZ HERNÁNDEZ (1997).

Il modello proposto non prevede tuttavia la semplice contrapposizione fra presenza e assenza di un tratto: non è possibile infatti intendere /maschile/ come assenza del tratto /femminile/ o, cambiando prospettiva, /femminile/ come assenza del tratto /maschile/. È da rilevare inoltre che questo genere di opposizione semantica è tanto produttivo, a livello di sistema funzionale della lingua, quanto lo schema basato sulla presenza o assenza di un tratto, come /luce/ contrapposto a /tenebra/⁷. Così, vedendo le aporie che un binarismo di impostazione jakobsoniana (marcato - non marcato) privo di ulteriori precisazioni avrebbe sollevato in semantica, Greimas introduce la nozione di asse semantico, «ce fond sur lequel se dégage l'articulation de la signification»⁸, e gli attribuisce una organizzazione simile a quella teorizzata, per altri scopi, da Brøndal⁹.

Nella sua completezza la categoria massimale brøndaliana, oltre a una coppia di termini disgiunti, complementari, comprende un termine neutro, che indica l'assenza delle qualità proprie dei termini disgiunti, un termine complesso che le sussume e due termini complessi sbilanciati rispettivamente verso l'uno o l'altro polo rappresentato da ciascun termine complementare. Un esempio classico è dato dalla categoria timica¹⁰, costituita da una coppia di opposti: *euforia - disforia*, dal termine complesso che li sussume, *diaforia*, dal termine neutro *adiaforia*, dal termine complesso positivo *non-disforia* e negativo *non-euforia*. Una rappresentazione della categoria timica appena vista è esemplificata in DUCROT E SCHAEFFER (1995, p. 282) per illustrare l'esagono logico di Blanché¹¹: *excité -*

⁷Serve appena ricordare come il problema sia ben noto alla filosofia antica.

⁸GREIMAS (1995, p. 21).

⁹Si veda DUCROT E SCHAEFFER (1995, pp. 279-280) per una esposizione sintetica del principio di partecipazione brøndaliano, mutuato dall'antropologo Lévy-Bruhl, secondo il quale tale principio caratterizzerebbe la mentalità primitiva.

¹⁰Si veda FABBRI E MARRONE (2000, p. 242).

¹¹Gli autori rilevano le analogie fra la categoria massimale di Brøndal e l'esagono logico di Blanché, tuttavia osservano che il termine che sussume i contrari in Blanché non comporta contraddizione, essendo definito come *A oppure B* (il ciclotimico è eccitato *oppure* depresso) mentre in Brøndal il termine complesso comporta contraddizione, essendo definito come *A e anche B*. È necessario ovviamente distinguere fra coppie di contrari e coppie di complementari: nella coppia di contrari B è definito come *non A*, e quindi la formula *A e B* comporta contraddizione logica. Nella coppia di complementari B è incompatibile con A, ma non è definito in

6. Modelli e risorse lessico-semantiche

déprimé sono la coppia di opposti, *cyclothymique* il termine complesso, *équilibré* il neutro, *en train* e *calme* rispettivamente il complesso positivo e negativo.

Nonostante questa articolazione dell'asse semantico, Greimas dichiara che «la structure brøndalienne est tout aussi binaire que celle de Jakobson»¹², e questo comporta che «le nombre de sèmes impliqués dans celle-ci [*scil.* l'axiomatique de Brøndal] reste constant». Ne consegue che l'analisi componenziale rimane nell'ambito del discreto, in quanto riconducibile a opposizioni binarie. Greimas non perde mai di vista il fatto che l'intima essenza della semantica strutturale è la sua natura differenziale¹³.

La semantica interpretativa

Sviluppando l'impianto teorico di Greimas, Rastier elabora una semantica particolarmente attenta alla pluralità dei sensi¹⁴. Tuttavia l'interesse non è rivolto ai sensi potenziali del testo, alle letture possibili, quanto piuttosto alle letture storicamente determinate, che si concretizzano in nuovi testi che fanno riferimento al testo in oggetto (sotto forma di commentario, di traduzione, di rifacimento, etc.):

Un texte ne peut pas avoir un nombre indéfini de sens. À un moment donné, il n'en a qu'un nombre fini et non clos, le sens s'égalant aux commentaires fixes. En multipliant les interprétations, la tradition culturelle les limite: le non fini n'est pas l'infini. Le sens est clos quand le texte n'est plus lu¹⁵.

Tutta protesa verso il testo¹⁶, la semantica interpretativa si fa forte dell'idea che due soli piani, *langue* e *parole*, non siano sufficienti a rendere ragione delle articolazioni effettive fra asse paradigmatico e asse sintagmatico: in effetti, l'osservazione che la dicotomia saussuriana non permetta di comprendere la totalità dei fatti linguistici reali, espressa con argomenti stringenti da Coseriu in termini di negazione di A. La loro congiunzione, A e B, non comporta quindi, dal punto di vista della logica formale, contraddizione.

¹²GREIMAS (1995, p. 24).

¹³Si veda RASTIER ET AL. (1994, pp. 31-34).

¹⁴Si veda nello specifico RASTIER (1996, pp. 167-212).

¹⁵*L'affaire du taxème rebondit!* - Échange entre François Rastier, Ioannis Kanellos et Régis Burnet (1995) <http://www.revue-texto.net/Dialogues/Taxeme.html>.

¹⁶Al contrario della semantica di Pottier e Greimas, che ragiona sul vocabolario.

*Sistema, norma y habla*¹⁷, è fra le acquisizioni più importanti della linguistica strutturale dopo la sua fondazione ginevrina. Infatti, se la selezione attuata a livello di *parole* venisse fatta attingendo direttamente al sistema funzionale della *langue*, senza mediazioni, ci si dovrebbe aspettare che i discorsi realizzati siano distribuzioni armoniose delle possibilità combinatorie offerte dal sistema stesso. Tuttavia non è così, e qualunque studio eseguito su un *corpus* rappresentativo¹⁸ di testi lo può dimostrare. Fra sistema funzionale e discorso¹⁹ Coseriu individua il piano intermedio della norma, che è definibile come il caso tradizionalmente più in uso di realizzazione del sistema. Dunque, fra la potenzialità della *langue* e l'attualizzazione della *parole* la norma si frappone come filtro, come fattore di selezione socialmente determinato. Come corollario ne consegue che il paradigma generativo, di matrice chomskiana, che vede il testo come il prodotto dell'applicazione di un numero relativamente ristretto di regole di composizione sulle unità di un lessico dato, non può rendere ragione dei testi storicamente esistenti. Rastier, molto attento ai rapporti fra semantica di scuola francese e tendenze di oltreoceano, su questo punto afferma con chiarezza, parafrasando de Saussure²⁰:

Risquons un truisme: rien ne peut être représenté en langue qui n'ait auparavant été décrit en contexte.

e precisa nella relativa nota la sua distanza dal generativismo:

Deux difficultés apparaissent ici. Si l'on refuse tout corpus déterminé, comme font les générativistes qui cherchent à engendrer les phrases *possibles*, on délaisse les textes réalisés pour s'attacher aux textes réalisables. Par là, on élimine *a priori* les normes à l'œuvre dans tout corpus déterminé; et l'on se retrouve dans l'impossibilité d'éliminer l'infinité des phrases grammaticalement possibles, mais indicibles parce qu'impensables: *Toute onde est associée à un corpuscule* était au XVII^e siècle une phrase possible, mais inconcevable.

¹⁷COSERIU (1973, pp. 11–113).

¹⁸La rappresentatività del *corpus* è data da una equa distribuzione diacronica e/o diatopica, diastratica, diafasica dei testi in esso compresi.

¹⁹Negli autori citati *discours* è da intendere come sinonimo del saussuriano *parole*.

²⁰RASTIER (1996, p. 62).

6. Modelli e risorse lessico-semantiche

Ma il quadro necessita per Rastier²¹ di un'ulteriore precisazione e mutua quindi da Hjelmslev la quadripartizione in *schéma*, *norme*, *usage*, *acte* per articolare su tre piani le istanze immanenti di codificazione, e destinare al fenomeno manifestato il piano dell'atto di parola²². Questi tre piani di codificazione quindi vengono a essere: il *sistema* funzionale della lingua, la *norma* determinata dall'orizzonte culturale della società e l'*uso* personale.

È attraverso il riconoscimento di questi piani intermedi, la norma e l'uso (con le dovute proiezioni sull'asse diacronico), che la linguistica post-saussuriana e la filologia hanno visto la possibilità di un riavvicinamento. Rispetto a una lingua senza parlanti nativi, quindi privata della chomskyana *competence*, il modo privilegiato per ricostruire (parzialmente) il codice è partire dalla totalità delle *performances* attestate, cioè dal *corpus*, come consegnato dalla tradizione. Tuttavia, se l'oggetto ricostruito per induzione fosse soltanto il sistema funzionale, cioè la lingua astratta intesa come lessico, regole di formazione delle parole, regole di combinazione delle unità lessicali, etc., questo oggetto risultante, questo sistema, non potrebbe essere utilizzato per valutare la qualità, la bontà o i guasti dei testi impiegati per costruirlo, se non limitatamente ai casi irrealizzabili per il sistema, cioè errori universalmente riconoscibili.

La norma e l'uso della linguistica acquistano quindi particolare interesse per la filologia, in quanto è soprattutto su questi piani che essa si trova a operare, in particolar modo per quanto concerne gli strumenti critici della *lectio difficilior* e dell'*usus scribendi*. Sono necessarie però delle precisazioni: soltanto un modello semplicistico sarebbe tentato di ritenere che l'*usus* possa essere ricostruito con i soli mezzi della linguistica statistica: la rilevanza statistica dei risultati di un processo di normalizzazione potrebbe falsare totalmente la percezione dell'*usus* genuino dell'autore: per la filologia, il concetto sincronico di *usus scribendi* è chiamato a confrontarsi con il concetto diacronico di *usus corrigendi*, che non

²¹RASTIER (1996, p. 40).

²²Benché Rastier dica esplicitamente di mutuare questa quadripartizione dagli *Essais linguistiques* di Hjelmslev, tuttavia una partizione fra norma sociale e norma individuale è già presente in COSERIU (1973, pp. 96–101). Per alcune precisazioni sulle nozioni di sistema, norma, uso nei due linguisti, si veda DUCROT E SCHAEFFER (1995, pp. 311–320).

coinvolge soltanto la dimensione idioletale del correttore, ma si estende alla dimensione sociolettale nella quale è attiva la norma. Anche la nozione di *difficilior* opera principalmente sui piani dell'uso e della norma: la valutazione del grado di difficoltà di una lezione coinvolge, diacronicamente, la ricostruzione della norma e dell'uso dell'autore e la ricostruzione almeno della norma, se non anche dell'uso, operante nel contesto in cui la variante si è prodotta.

Se il discorso vale a livello di costituzione del testo, è ancor più valido a livello di interpretazione, e questo è il motivo per cui la semantica di Rastier, forte della speculazione linguistica della seconda metà del secolo scorso, può essere di grande stimolo negli studi filologici. L'autore, che si è occupato del rapporto della semantica strutturalista con l'informatica²³ e, più nello specifico, del rapporto (anche conflittuale) con l'Intelligenza Artificiale di impostazione cognitivista²⁴, attraverso l'accettazione nel quadro interpretativo generale della nozione di norma (come intesa da Coseriu), mette in guardia il linguista informatico dalla creazione di modelli troppo semplicistici. Come si è appena visto, la tentazione più pericolosa è ritenere che le sole informazioni di carattere linguistico immanenti ai testi di un *corpus* redatto su edizioni di riferimento siano sufficienti a ricostruire l'uso e il sistema funzionale. La comprensione della norma, che filtra il «grammaticalmente possibile» dando ragione dell'«indicibile perché impensabile» non coinvolge soltanto la competenza linguistica, ma anche l'orizzonte etico-religioso, le conoscenze scientifiche (come nell'esempio sulla natura ondulatoria e corpuscolare della luce), il diritto, etc.: in una parola la *civilisation*²⁵ della società che la produce.

Secondo Rastier, quindi, la descrizione semantica di un testo che tenga conto soltanto dei tratti inerenti, pertinenti al sistema funzionale della *langue*, è senza dubbio insufficiente. Una descrizione adeguata deve comprendere anche tratti semantici afferenti, la cui afferenza risponda a una norma socialmente determinata e, per finire, tratti semantici idiolettali.

²³Si veda RASTIER ET AL. (1994).

²⁴RASTIER (2001).

²⁵Si veda RASTIER (1996, p. 63).

6. Modelli e risorse lessico-semantiche

Concetti e terminologia della semantica interpretativa

Siccome la terminologia usata da Pottier, Coseriu, Greimas e dalla scuola francese di semiotica ha subito piccole variazioni nel corso del tempo e si è arricchita di nuovi guadagni speculativi, sembra qui opportuno illustrare il quadro concettuale e terminologico, così come è utilizzato nella semantica interpretativa di Rastier e dei suoi seguaci, che noi adottiamo nel presente lavoro²⁶.

Il *morfema* è il segno linguistico minimo, il quale può appartenere a una o più classi aperte, in uno stato sincronico dato, assumendo il nome di *lessema*, oppure a una o più classi chiuse, assumendo il nome di *grammema*. Il morfema può essere tanto libero quanto legato: nel primo caso corrisponde a una unità lessicale, a una voce di vocabolario, nel secondo caso si trova stabilmente in composizione con altri morfemi. Per questo un grammema può rappresentare tanto una parte funzionale del discorso, come le preposizioni e le congiunzioni, quanto un preverbo o una desinenza. La precisazione, nella definizione, di stato sincronico, dipende dal fatto che, in prospettiva diacronica, un lessema può andare ad assumere stabilmente la funzione di grammema: è il caso ad esempio di avverbi che si sono stabilizzati in funzione di preposizioni improprie. Data la definizione, i lessemi possono invece trovarsi soltanto, liberi o in composizione con altri lessemi o grammemi, in unità appartenenti alle classi aperte dei sostantivi, aggettivi, verbi e avverbi. Un raggruppamento stabile di morfemi costituente una unità funzionale assume il nome di *lessia*. Ciascuna forma flessa o indeclinabile riconducibile a una voce di dizionario rappresenta una di queste unità, ma il concetto di lessia non risponde alla definizione di parola (per il testo scritto) come raggruppamento di grafemi separati da spazi bianchi. La lessia, in quanto unità funzionale, può essere costituita da più parole: la locuzione avverbiale ne è un esempio.

²⁶Queste precisazioni potrebbero sembrare superflue, visto che sono facilmente reperibili consultando i glossari in appendice alle opere citate di Rastier oppure nei glossari presenti sul sito curato dall'autore: <http://www.revue-texto.net> (dove si trova, fra l'altro, una ricca bibliografia di semantica testuale). Tuttavia si è deciso di riportare, commentando, i termini del glossario per evitare fraintendimenti dovuti all'uso dei medesimi termini da parte di autori diversi: un esempio fra tutti, l'uso nella *Sintassi generale* di Martinet del termine monema al posto di morfema e di morfema al posto di grammema.

In quanto segno, il morfema è dotato di un significante, di cui per definizione la semantica non si occupa²⁷, e di un significato, detto *semema*; il significato di una lessia è detto invece *semia*. Il semema può essere suddiviso in unità minimali di significato, ciascuna delle quali è detta *sema*²⁸. Data la natura differenziale della semantica interpretativa e, più in generale, della semantica strutturale, i semi sussistono soltanto in un sistema di opposizioni fra sememi, tanto che Rastier, nei suoi glossari²⁹, definisce il sema come l'estremità di una relazione funzionale binaria tra sememi. Le conseguenze sono della massima importanza e non sono state sufficientemente prese in considerazione, almeno in sede di impostazione teorica, magari per prenderne le distanze, dalla linguistica computazionale di scuola americana, come si vedrà più avanti. La prima conseguenza riguarda il fatto che i semi non sono degli universali linguistici: per il fatto di essere tratti distintivi di sememi che rappresentano il contenuto di morfemi appartenenti, sull'asse paradigmatico, a lingue storicamente attestate e sull'asse sintagmatico a testi effettivamente realizzati, tali tratti distintivi fra coppie o tuple di sememi non possono essere universali, in quanto determinati all'interno di un particolare sistema funzionale, norma o uso e non universalmente applicabili, aprioristicamente, a qualsiasi testo espresso in qualsiasi lingua. I semi non sono neppure componenti ultime o primitive, avendo valore relativo di opposizione ed essendo invece privi di valore ontologico o cognitivo. Deve quindi essere tenuto distinto il sema dal *noema*, componente primitiva la cui esistenza è decisa dall'analista indipendentemente dalla o dalle lingue naturali analizzate: il noema è quindi un universale di metodo, non un universale empirico e il suo valore è soltanto operativo. Per finire, il numero dei semi è stabilito dal numero di opposizioni che

²⁷Per una descrizione sul piano del significante del morfema (monema) si può far riferimento a MARTINET (1985), dove sono trattati per esteso, da un punto di vista funzionalista, anche i concetti di grado zero e amalgama.

²⁸Per quanto nella lingua italiana si possa tradurre questo termine con *tratto semantico*, si è preferito qui, in sede di definizione, rendere il francese *sème/sèmes* con *sema/semi*, in quanto nelle opere di Rastier il termine *trait sémantique* è utilizzato soprattutto per rendere il termine *semantic feature* di scuola americana oppure come termine non tecnico. Tuttavia, nel corso del presente lavoro, ci capiterà di impiegare *sema* e *tratto semantico* come sinonimi.

²⁹Tali glossari sono in appendice a tutte le opere citate dell'autore.

6. Modelli e risorse lessico-semantiche

essi devono determinare fra sememi e poiché la combinatoria di questi ultimi non è libera, anche quella dei semi non lo può essere e quindi il loro numero non è direttamente funzione del numero di semi che vanno a differenziare. Rastier³⁰ afferma che l'esistenza di strutture paradigmatiche del lessico impedisce tuttavia che il numero dei semi sia troppo grande, poiché i sememi non sono interdefiniti tutti insieme (ma all'interno di paradigmi semantici), e impedisce anche che tale numero sia troppo piccolo, proprio perché la combinatoria dei sememi non è libera.

Il lessico può essere strutturato, a livello di sistema, di norma e di uso, in paradigmi semantici attraverso il riconoscimento, all'interno dei sememi, di *semi generici*, che determinano l'appartenenza del semema a una medesima classe e di *semi specifici*, che lo differenziano dagli altri membri della classe. L'insieme dei semi generici prende il nome di *classema*, mentre l'insieme dei semi specifici prende il nome di *semantema*. Classema e semantema costituiscono quindi due sottoinsiemi dell'insieme non ordinato di semi che forma il semema. Mentre i tratti specifici appartenenti al semantema non sono organizzabili in strutture intermedie, i tratti generici possono essere organizzati su tre diversi ordini di grandezza; si vengono quindi ad avere semi *macrogenerici*, *mesogenerici* e *microgenerici*. La loro esistenza è data nel sistema funzionale della lingua, mentre la loro attualizzazione o virtualizzazione, come si vedrà in seguito, ha luogo in contesto.

La dimensione

A livello macrogenerico, i sememi si strutturano in *dimensioni*, che possono essere viste come insiemi di tratti semantici di estensione massima, raggruppati in un numero molto ristretto (sull'ordine delle decine) di categorie chiuse: //animato// - //inanimato//, //maschile// - //femminile//, //concreto// - //astratto//, //perfettivo// - //imperfettivo// ne sono esempi. Diacronicamente, le dimensioni variano la loro organizzazione in tempi molto lunghi, millenari, ed è per questa ragione che esse generalmente non sono soltanto lessicalizzate, ma sono pure gram-

³⁰RASTIER (1996, p. 32).

maturalizzate. Così, l'accordo non istituisce solo il legame sintattico fra termini appartenenti a categorie grammaticali diverse ma segnala anche un certo tipo di legame semantico (istituisce cioè una isotopia³¹ macrogenerica) fra di loro.

Il dominio

A livello mesogenerico, i tassemi, cioè le classi minimali di sememi, possono essere organizzati in *domini*, all'interno dei quali non si dà polisemia. Il fatto che nella nozione di dominio rientri la disambiguazione dei casi di polisemia necessita di un chiarimento, dato che tutta la teoria di strutturazione dei sememi fin qui vista si articola sul piano del contenuto e non sul piano dell'espressione, mentre la polisemia implica, per definizione, che sia un significante (piano dell'espressione) ad avere molteplici significati (piano del contenuto). Come è noto, la prospettiva della semantica strutturale è onomasiologica, differenziandosi in questo dalla prospettiva semasiologica della lessicografia e della semantica di scuola americana, tanto generativista quanto cognitivista:

L'approche différentielle, illustrée notamment par Pottier et Coseriu, adopte une méthode *onomasiologique* : les sémèmes sont analysés par leur différences au sein d'une classe sémantique minimale (le taxème), constituée par un critère strictement sémantique [...]

L'approche référentielle, illustrée notamment par Katz s'en tient à la méthode *semasiologique* caractéristique de la lexicographie: les sémèmes sont analysés par leurs différences au sein d'une classe constituée par un critère d'*expression*³².

In effetti, proprio Rastier ammette che la nozione di dominio è familiare alla lessicografia: «tous les dictionnaires recourent au moins implicitement au concept de domaine, en utilisant des abréviations qui signalent l'appartenance de contenus à des domaines comme *cuis.* (terme de cuisine) [...] ou *mar.* (terme de marine) pour *carguer*.» La costituzione e la struttura interna dei domini è legata a norme

³¹Per la nozione di isotopia si veda *infra*.

³²RASTIER (2001, p. 141). Si veda anche, fra gli altri, MARTÍNEZ HERNÁNDEZ (2000, p. 15), che dice cursoriamente: «mientras que el campo semasiológico se funda en la polisemia, el campo onomasiológico lo hace en la sinonimia».

6. Modelli e risorse lessico-semantiche

sociali, in particolare all'organizzazione delle attività umane e alla specializzazione dei saperi e delle tecniche. È all'interno dei domini che si sviluppano i lessici specifici, materia di studio non solo della lessicologia ma anche della terminologia³³. Se le dimensioni, come si è visto, subiscono modifiche in tempi millenari (ad esempio, la soppressione di una modalità verbale) in quanto agiscono profondamente sul sistema funzionale della lingua, non solo a livello semantico ma anche morfosintattico, i domini invece possono subire riorganizzazioni in tempi secolari o in tempi ancora più rapidi, poiché seguono il ritmo dell'organizzazione delle attività e delle acquisizioni nei campi del sapere di cui sono espressione. Ad esempio, l'evoluzione della scienza medica in età classica comporta, nella lingua, un conseguente cambiamento del dominio semantico della medicina.

Il concetto di dominio, proprio per il suo legame stretto con l'organizzazione sociale, extralinguistica, è entrato relativamente tardi nel paradigma della semantica differenziale, fortemente orientata alla costruzione del sistema funzionale della lingua. È stato necessario comprendere fino in fondo che la nozione di dominio non è un criterio di classificazione dei referenti, dei *realia*³⁴, bensì di organizzazione del contenuto linguistico. Con tutta la prudenza del caso e con le dovute distinzioni, ci sembra che la nozione di dominio della semantica interpretativa abbia punti di contatto con la nozione di *istituzione* proposta da BENVENISTE (1969, pp. 9–11), benché la prospettiva del *vocabulaire* sia diacronica (e diatopica) e non sincronica (e sintopica) come quella della teoria semantica di cui ci si occupa qui:

Le terme d'institution est à entendre ici dans un sens étendu: non seulement les institutions classiques du droit, du gouvernement, de la religion, mais aussi celles, moins apparentes, qui se dessinent dans les techniques, les modes de vie, les rapports sociaux, les procès de parole et de pensée. [...] On s'efforce ainsi [...] de montrer comment les langues réorganisent leurs systèmes de distinctions et rénovent leur appareil sémantique.

L'aspect historique et sociologique de ces procès est laissé à d'autres. [...]

³³Si veda MAGRIS ET AL. (2002).

³⁴Il dominio inteso come criterio di classificazione di *realia* è ben presente in tutta la tradizione enciclopedica, antica e moderna. Gli *homerische Realien* di Buchholz sono uno dei più noti esempi di organizzazione del lessico di un autore in funzione della ricostruzione dei *designata*.

Nous éclairons par là la *signification*; d'autres se chargeront de la *désignation*.
[...] Les historiens et les sociologues verront mieux alors ce qu'ils peuvent retenir des présentes analyses où n'entre aucun présupposé extralinguistique.

Ma la nozione di dominio (o, per usare le sue parole, di «istituzione in senso esteso») gli permette di valutare la plausibilità o meno della genesi comune di radici compatibili dal punto di vista del significante ma appartenenti a domini di significato lontani: parlando del latino *texo* e del sanscrito *takṣ* (rispettivamente “tessere” e “tagliare con l'ascia”), Benveniste nota:

on ne voit pas comment un de ces sens pourraient être dérivé de l'autre ni de quel sens antérieur ils pourraient l'un et l'autre procéder, “tissage” et “charpentage” paraissant irréductibles à une technique commune.

La nozione di dominio, importante sul piano paradigmatico per rendere ragione della polisemia, sul piano sintagmatico permette di classificare similitudini e metafore che associano fra di loro ambiti diversi dell'esperienza e dell'attività umana: VAN NES (1963) ad esempio si è concentrato sullo studio di un singolo dominio, quello dei termini marittimi, dimostrando quanto sia produttivo nella poetica di Eschilo per la creazione di immagini.

Ed è proprio a livello mesogenerico che si attua la distinzione fra catacresi, o metafora morta e metafora viva. Come nota Roussel, in *Pers.* 71-2 Eschilo offre uno dei numerosi esempi della sua capacità di rivitalizzare una metafora morta:

*πολύγομφον ὄδισμα
ζυγὸν ἀμφιβαλὼν ἀχένη πόντου.*

L'operazione è raffinata, perché *ἀχίην*, nel dominio dei termini geografici, ha il significato di “stretto” (detto di un braccio di mare) o di “punto di restringimento” (detto di fiume)³⁵. Del resto, anche il verbo *ζεύγνυμι* ha il significato specifico di “congiungere due rive con un ponte”³⁶. Ma paragonando la «via dai

³⁵In Erodoto il termine è usato parlando del Bosforo (Hdt. 4.85), del Danubio (Hdt. 4.89), etc. Come termine geografico indica anche “istmo” (Hdt. 1.72, 6.37) o “passo montano” (Hdt. 7.223).

³⁶Fra i molti esempi, Hdt. 1.206.

6. Modelli e risorse lessico-semantiche

molti chiodi» a un giogo, istituendo quindi una relazione di similitudine fra domini differenti, l'Ellesponto viene personificato, così *αὐχένι (πόντου)* torna a essere una metafora viva. Roussel commenta: «Il y a donc ici une image renouvelée; c'est encore mieux qu'une image nouvelle».

Differenza fra le nozioni di dominio e di script

Per concludere, è opportuno precisare la differenza fra dominio³⁷ e *script*: nozione questa che, sviluppatasi in seno agli studi sull'Intelligenza Artificiale di matrice cognitivista³⁸, si sta diffondendo negli studi di letteratura classica³⁹.

La nozione di dominio ha un valore prettamente linguistico e stabilisce il criterio per individuare differenti nomenclature: lo si può intendere quindi come una collezione non ordinata di termini. Al contrario il concetto di *script* esprime una situazione socialmente codificata. Il suo valore primario è extralinguistico, in quanto descrive, indipendentemente dal linguaggio, le procedure o gli stati di cose attesi. Seguendo il paradigma cognitivista, sul quale si tornerà in seguito, uno *script* può essere più o meno tipico in base alla sua somiglianza rispetto a un prototipo. La somiglianza tuttavia non è stabilita in base a un nucleo comune di tratti identici, ma rispetto a un'«aria di famiglia»⁴⁰ che permette di accomunare tre o più elementi in base a tratti condivisi fra coppie di elementi, anche quando nessun tratto sia condiviso dall'intero insieme.

Perché uno *script* sia identificabile deve ripetersi almeno due volte e perché si possa riconoscere il suo prototipo, quando non si faccia ricorso a una petizione di principio, deve avere almeno tre occorrenze, due delle quali più simili fra di loro rispetto alla terza. In Omero quasi ogni sequenza di azioni, dalla vestizione alla rassegna delle armi, dal duello alla preghiera è codificabile in *script*.

³⁷Si veda, per questa distinzione, RASTIER (1996, p. 49 n. 18).

³⁸Si veda SCHANK E ABELSON (1977).

³⁹Un esempio di applicazione esplicita nell'ambito degli studi omerici è il volume della MINCHIN (2001). Tuttavia il concetto di *script* trova un parallelo indipendente nell'impostazione di PAVESE (1968), che vede il suo compimento in PAVESE (1997).

⁴⁰L'espressione è esplicitamente mutuata, da parte dei cognitivisti, dalle *Ricerche filosofiche* di Wittgenstein.

Del resto, per sua stessa definizione, un rito ha la natura di uno *script*. Nell'evocazione di Dario (*Pers.* 607-680) è in atto uno *script* che si articola in due parti: l'una vicina al prototipo iliadico (le *χοαί* della regina Atossa), l'altra più innovativa (gli *ῥμνοι* agli dei inferi). Le libagioni della regina offrono una serie completa di latte (611), miele (612), acqua (613), vino (614-15), olio⁴¹ (616-17) e fiori (618). Nella *Nekyia* (λ 27-28, anticipati in κ 519-20):

πρῶτα μελικρήτω, μετέπειτα δὲ ἡδέϊ οἴνω
τὸ τρίτον αἶθ' ὕδατι ἐπὶ δ' ἄλφιστα λευκὰ πάλυνον.

troviamo una mistura di miele e latte⁴², vino e acqua, più uno spargimento di farina. Come ricorda Belloni, nel suo commentario *ad loc.*, le offerte di Atossa trovano parziali riscontri in altri luoghi dell'epica e della tragedia, con variazioni rispetto al prototipo funzionali al contesto. In *Ψ* 220 Achille versa il vino per l'ombra di Patroclo⁴³; Oreste depone sulla tomba di Agamennone latte, fiori e una ciocca di capelli in Soph. *El.* 894-6; Ifigenia dedica a Oreste latte, vino e miele in Eur. *IT* 157-66; Ermione destina a Clitemnestra miele misto a latte e vino in Eur. *Or.* 114-16; Edipo offre acqua e miele escludendo esplicitamente il vino in Soph. *OC* 481, così come il vino è escluso in *Eum.* 107. Come si vede, latte e miele rappresentano i prototipi di queste offerte mentre la ciocca di capelli tradisce la sua atipicità, funzionale al motivo dell'agnizione. Si è visto dunque che la prima parte dello *script*, la libagione, nei *Persiani* è tutta greca (con l'ampliamento di olio e fiori, comunque possibile), e questo permette di mettere ancor più in risalto l'unicità della seconda parte, dove, dopo l'invocazione alle divinità ctonie ci si rivolge, con le forme dell'inno, a un mortale. Nota Citti⁴⁴:

⁴¹Propriamente *ἐλαίας καρπός*, ma la metonimia del frutto per il suo prodotto ci è confermata ad esempio da *Γ* 246, dove *καρπὸν ἀρούρης* è apposizione di *οἶνον*.

⁴²Per il fatto che *μελικρήτω* sia da intendere come una mistura con latte, si veda EBELING (1885), *s.v.* In Eur. *Or.* 115 il secondo componente è detto in modo esplicito: *μελίκρατα γάλακτος*, tuttavia in altri contesti il termine può indicare una mistura con acqua: *e.g.* Hp. *Aph.* 5.41.

⁴³Belloni cita anche *Ψ* 170, dove l'eroe appresta *μέλιτος καὶ ἀλείφατος ἀμφιφορηῆς*, tuttavia queste offerte sono arse nel rogo e non versate a terra: difficilmente si possono considerare i vv. 170 e 220 come una triade di miele, olio e vino per un rito assimilabile al nostro in oggetto.

⁴⁴CITTI (1962, pp. 41-43).

6. Modelli e risorse lessico-semantiche

Ma la preghiera vera e propria è più oltre, e non è agli dei inferi, bensì al *Περσῶν Σουσιγενῆ θεόν*, al re morto, in modo che la strofe e l'antistrofe γ assumono il tono dell'inno al dio di cui si attende l'epifania.

L'atipicità è amplificata anche dal fatto che

non vi sono nei *Persiani*, a quanto ho potuto vedere, altri brani di derivazione sacrale o liturgica, se tralasciamo le lamentazioni funebri, che meriterebbero un discorso a parte.

All'interno di un medesimo *script* possono essere impiegati termini appartenenti a domini differenti, come il dominio delle offerte rituali e il dominio liturgico.

Il tassema

Fatte queste precisazioni, si può ora prendere in considerazione lo stadio più basso di raggruppamento dei tratti semantici. A livello microgenerico, i sememi sono organizzati in classi minimali dette *tassemi*, le quali, per definizione, non ammettono intersezioni reciproche. È all'interno del tassema che i sememi si interdefiniscono: i sememi microgenerici comuni stabiliscono il criterio di appartenenza alla classe stessa, mentre i sememi specifici li distinguono e li identificano gli uni rispetto agli altri. La nozione di tassema permette di uscire dall'*impasse* todoroviana secondo cui, dato che il significato è determinato all'interno di un sistema di opposizioni, allora devono essere stabiliti tanti tratti differenziali quante sono le combinazioni possibili fra tutti i sememi della lingua. Poiché si interdefiniscono all'interno del tassema, il numero di tratti sufficienti a differenziarli, e quindi il numero di sememi specifici appartenenti al semantema, è relativamente ridotto. L'opposizione dei termini all'interno di un tassema può acquistare un valore emblematico, o per metafora, o per metonimia. Un esempio del primo caso è rappresentato dal tassema dei rapaci nel racconto di Atossa, dove aveva visto l' *αἰετὸν* del v. 205, che per maestosità rappresenta il Persiano, sopraffatto dal *κίρκον* del v. 207. Un esempio del secondo è rappresentato dal tassema delle armi, che rappresentano metonimicamente il possessore attraverso l'arma caratterizzante: 147-149 *τόξου (ρῦμα)* dei Persiani contro *λόγχης (ἰσχὺς)* dei Greci.

È da notare che il tassema ha valore *en langue*, tuttavia la sua esistenza può essere provata soltanto *en discours*: il processo interpretativo, che caratterizza la semantica di Rastier, può partire soltanto dal realizzato verso il realizzabile, e non viceversa, come propone il paradigma generativo. Il tassema quindi non si costruisce a priori, non è un universale di metodo come il noema, ma si individua attraverso lo studio di *corpora*. Sono in particolare gli enunciati storicamente attestati che contengono disgiunzioni (e.g. *Pers.* 248 ἐσθλὸν ἢ κακόν), congiunzioni (e.g. *Ag.* 5 χεῖμα καὶ θέρος) o altri sintagmi in cui sia riconoscibile una struttura logica forte che permetta di identificare i tassemi (rispettivamente, per gli esempi visti, di //giudizio di valore// e di //partizione dell'anno//).

Si è scelto intenzionalmente l'esempio di *Ag.* 5 perché permette di fare alcune osservazioni sulla nozione di interdefinizione. Si è detto che il tassema ha valore *en langue*, quindi, in primo luogo, una singola cooccorrenza *en discours* non dà alcuna garanzia che una certa collezione di sememi costituisca di fatto un paradigma semantico appartenente al sistema funzionale della lingua. Come è noto⁴⁵, il numero delle stagioni in Grecia ha variazioni diacroniche: dalle tre di Omero ed Esiodo, alla comparsa di una quarta (ὀπώρα), fluttuante in età classica, alla menzione di sette in età ellenistica. Ha inoltre, presumibilmente, variazioni diatopiche: *Alcm.* 76 (Bergk): θέρος καὶ χεῖμα κὼπώραν τρίταν καὶ τέτρατον τὸ Fῆρ, testimonierebbe, in ambiente dorico, la quadripartizione. Per finire, si può pensare anche a variazioni diastratiche: quattro o tre ai fini dell'organizzazione di attività agricole o di altre attività civili, due, inverno ed estate, ai fini dell'organizzazione di attività militari (come ci attesta la suddivisione tucididea del materiale trattato).

Nel commentario a *Prom.* 454-6

ἦν δ' οὐδὲν αὐτοῖς οὔτε χεῖματος τέκμαρ
οὔτ' ἀνθεμῶδους ἦρος οὔτε καρπίμου
θέρους βέβαιον, ἀλλ' ἄτερ γνώμης τὸ πᾶν

Griffith, *ad loc.*, dice:

⁴⁵Per una panoramica generale, senza approfondire ulteriormente, si rimanda al Daremberg-Saglio, s.v. *Horae*.

6. Modelli e risorse lessico-semantiche

The Greeks tended to think in terms of three, rather than four, seasons (cf. Diod. Sic. 1.26.5⁴⁶, LSJ *s.v.* ὀπώρα, N.J. Richardson's n. on *Hom. Hymn Dem.* 399ff.).

Un'ulteriore precisazione viene fornita da Fraenkel, nel commentario ad *Ag.* 5:

It has been pointed out by Martin P. Nilsson, *Die Entstehung und religiöse Bedeutung des griech. Kalenders*, Lund 1918, 22, that the division of the year into two seasons was not abandoned by the Greeks even when other seasons had been added. In *PV* 454 ff. the beginnings of winter, spring, and summer are indicated by the stars.

Tuttavia sembra che i due commentatori non traggano le conclusioni dall'intero quadro delineato, cioè che nei due passi ci si trova di proposito di fronte a due paradigmi semantici differenti, ciascuno appropriato al contesto e quindi scelto consapevolmente dall'autore, in quanto nel *Prometeo* la divisione del tempo è data dal dio agli uomini come nozione astronomica funzionale alla coltivazione della terra, mentre nell'*Agamennone* la scelta scandisce il tempo secondo i ritmi binari della bella e della cattiva stagione, funzionali all'organizzazione di attività militari o di guardia. Se questa interpretazione è corretta, la differenza fra i due paradigmi semantici assume quindi una valenza diastratica.

Concludendo, questa breve digressione ci ha fornito un esempio di come si articola l'interdefinizione all'interno del tasema: *χειμα* e *θέρπος*, pur essendo in entrambe le occorrenze partizioni dell'anno, nel *Prometeo* sono tripartizioni (essendo menzionata anche la primavera al v. 455), nell'*Agamennone* invece sono bipartizioni dell'anno.

Semi inerenti e afferenti

Isolato dal contesto, un semema è costituito da un fascio di semi inerenti, vale a dire codificati *en langue* e quindi facenti parte della sua definizione nel vo-

⁴⁶Nel passo citato Diodoro Siculo ricorda che la tripartizione è importata dall'Egitto.

cabolario⁴⁷. Tuttavia nel contesto un semema può caricarsi di semi afferenti⁴⁸. Deve essere il processo interpretativo a rilevare quali meccanismi siano presenti nel contesto: da un lato infatti i semi inerenti possono essere virtualizzati fino a essere neutralizzati, dall'altro i semi afferenti possono essere attualizzati fino a diventare tratti salienti. Quando l'afferenza sia socialmente codificata e rientri quindi nella norma, essa dà origine ad accezioni diverse di uno stesso significato: in *Pers.* 133-4

λέκτρα δ'ἀνδρῶν πόθῳ
πίμπλαται δακρύμασιν

ἀνδρῶν si può caricare del tratto afferente di /marito/, appartenente al tassema //relazioni matrimoniali//, indotto dal contesto.

Ma il meccanismo di attualizzazione e virtualizzazione entra in gioco anche per descrivere gli usi figurati: in *Pers.* 169 ὄμμα γὰρ δόμων νομίζω δεσπότης παρουσίαν il tratto /parte del corpo/, inerente a ὄμμα, viene neutralizzato, mentre diventa saliente il tratto afferente di /parte migliore/.

L'isotopia

Il concetto di isotopia assume un ruolo centrale nella semantica di scuola francese, in quanto riguarda i meccanismi di selezione dall'asse paradigmatico degli elementi da comporre sull'asse sintagmatico. Nel discorso, l'isotopia⁴⁹ esprime la

⁴⁷La definizione di sema inerente è stata modificata da RASTIER (1996, p. 277): «extrémité d'une relation symétrique entre deux sémèmes appartenent à un même taxème» a RASTIER (2001, p. 258): «sème que l'occurrence hérite du type, par défaut. Ex.: /noir/ pour 'courbeau'». Benché Rastier critichi il concetto di prototipo caro al cognitivismo americano, con questa seconda definizione sembra tradire un forte avvicinamento, assimilando la nozione di *default*. In termini aristotelici, sarebbero tratti inerenti non soltanto le nozioni necessarie alla definizione, ma anche il *proprio*: è *proprio* del corvo essere nero, per quanto questa proprietà non rientri nella definizione di 'corvo'.

⁴⁸RASTIER (2001, p. 258) fa l'esempio di «/faiblesse/ pour 'femme'»: si vedrà nel commentario a *Pers.* 135 al Cap. 8 come proprio la presunzione di questa afferenza abbia spinto la maggior parte dei traduttori e dei commentatori a interpretare il passo.

⁴⁹Si veda GREIMAS (1995, p. 53).

6. Modelli e risorse lessico-semantici

ripetizione di tratti semantici generici che permette di interpretare correttamente un messaggio, accettando uno o più sensi plausibili a esclusione di tutti gli altri. Già il solo accordo di genere e numero istituisce un'isotopia minimale. Ma un'isotopia può estendersi per un intero testo: come è stato più volte rilevato, i *Persiani* sono costruiti sull'isotopia /grande quantità/.

Prendiamo ora l'esempio stesso di Greimas, dato che trova un curioso parallelo in uno dei vari tentativi di emendare il nostro testo eschileo. Si considerino quindi le frasi: «le chien aboie» e «le commissaire aboie». Secondo Greimas, a seconda dei contesti, «aboie» deve essere scomposto in tratti semantici differenti: il primo in /cri/ + /animal/, il secondo in /cri/ + /humain/, proprio perché la prima occorrenza istituisce un'isotopia con «chien», tramite la condivisione del tratto /animal/, mentre la seconda occorrenza istituisce un'isotopia con «commissaire» tramite la condivisione del tratto /humain/⁵⁰.

La presunzione d'isotopia agisce, almeno implicitamente, non soltanto in sede di interpretazione del testo ma anche di *constitutio*: il sospetto di una imperinenza semantica può spingere infatti a sanare il testo ristabilendo la presunta isotopia perduta. Per esempio, al tormentato v. 13 Rogers⁵¹ propone di emendare

⁵⁰Secondo la teoria di Rastier, il tratto /humain/ in «aboie» si può intendere come afferente, indotto dal contesto stesso; l'attualizzazione di tale tratto implica la neutralizzazione del tratto /animal/.

⁵¹Si veda DAWE (1965, *ad loc.*), CHANTRAINE (1999, *s.v.* βαύζω), citando MAZON (1950), precisa la costruzione sintattica e la valenza semantica del verbo: «complément à l'acc. [...], cet accusatif désignant la personne dont l'arrivée fait gronder un chien ou une personne».

νέον δ' ἄνδρα βαύζει con *κύων τ' ἄνδρα βαύζει*⁵², attratto dall'isotopia che verrebbe a stabilirsi fra *κύων* e *βαύζει*, verbo derivato dall'onomatopea *βαὺ βαύ*.

Di particolare interesse è la cooccorrenza di isotopie⁵³, in quanto permette di studiare l'*usus* sul piano del significato, a prescindere dalla cooccorrenza di vocaboli.

Isotopia e uso figurato

Come si è visto, il concetto di isotopia si basa sull'idea che le unità di significato (i sememi) che si dispongono a formare un discorso compiuto siano scomponibili in tratti semantici e che un certo numero di questi tratti siano condivisi da due o più unità, in modo tale che il discorso, in virtù di questa ridondanza semantica, trovi la sua coesione e coerenza. Tuttavia il linguaggio poetico e, più in generale, ogni forma di espressione figurata gioca sulla possibilità di forzare le isotopie attese nel linguaggio ordinario, attraverso meccanismi di attualizzazione e di virtualizzazione dei semi inerenti e afferenti. Negli anni successivi a *Sémantique structurale* si è sviluppata così una teoria delle figure di significato e di discorso⁵⁴ volta allo studio dei rapporti fra i tratti semantici del *designans* e i tratti del *designatum*.

⁵²Questi i vv. 10-13 nell'edizione di West:

*κακόμαντις ἄγαν ὀρσολοπέϊται θυμὸς ἔσοθεν
πᾶσα γὰρ ἰσχὺς Ἀσιατογενῆς
οἴχωκε, νέον δ' ἄνδρα βαύζει*

Per l'intelligenza del passo e per le differenti interpretazioni che ne sono state date, si rimanda a Belloni, *ad loc.* Qui basti ricordare che, a partire dagli scoli stessi, la critica si è divisa fra chi considera come soggetto di *βαύζει* il *θυμός* del v. 11, considerando *πᾶσα – οἴχωκε* una parentetica, e chi al contrario considera come soggetto di *βαύζει* *πᾶσα ἰσχὺς Ἀσιατογενῆς*, coordinando le due frasi contigue. Sia in un caso che nell'altro, la congettura di Rogers può essere intesa come predicativo del soggetto: «come un cane».

⁵³Dell'argomento si occupa soprattutto GROUPE μ (1990).

⁵⁴Si vedano in particolare MORIER (1981), GROUPE μ (1982), BREMOND (1994). Mentre l'intento di LAUSBERG (1990) è di descrivere la natura delle diverse figure partendo dalla trattatistica antica, l'impostazione strutturalista della *Rhétorique générale* non manifesta particolare interesse per la trattatistica antica e ridefinisce i termini di sineddoche, metafora, metonimia etc. per mezzo della teoria degli insiemi e della logica matematica.

6. Modelli e risorse lessico-semantiche

Questa impostazione ha prodotto criteri rigorosi per rendere espliciti e coerenti all'interno di un sistema omogeneo (e quindi, ai fini del nostro studio, codificabili mediante formule logiche) i meccanismi soggiacenti al riconoscimento di figure.

L'apporto maggiore di *Sémantique interprétative* è la consapevolezza che la figura non è equivalente a nessuna delle sue descrizioni, non può essere sostituita da nessuna di esse: se la figura appartiene al testo, la descrizione appartiene alla lettura. Nel linguaggio di Rastier, la «lettura» è un testo a vocazione metalinguistica: un commentario, una traduzione, etc.: è dunque un *altro* testo e come tale, ricorsivamente, si presta a essere oggetto di lettura. L'osservazione di TRUJILLO (2000, pp. 138-139) assume addirittura toni polemici:

No hay mayor error que el de dar un texto como significado de otro texto: en eso consiste, precisamente, esa práctica, con pretensiones científicas, del “comentario de textos”, tan en boga aún. [...]

Cuando reflexionamos sobre un texto, no hacemos otra cosa -nos guste o no- que crear un nuevo texto, o, lo que es lo mismo, que hacer literatura, mala o buena. En realidad, no hay crítica científica, propiamente hablando: la crítica es [...] creación, arte literario, y del arte literario lo único que puede predicarse es la cualidad.

Naturalmente, anche una descrizione formale di identificazione di tratti semantici, di codifica delle figure retoriche etc. non può che essere una lettura e, se si applica il metodo filologico espresso in MARTÍNEZ HERNÁNDEZ (1997), essa sarà una “lettura del testo” integrata da una “lettura di letture del testo”.

Un esempio per concludere: *Pers.* 51 *Μάρδων, Θάρυβις, λόγχης ἄκμονες*. L'uso figurato di *ἄκμονες* è garantito dalla rottura dell'isotopia basata sul tratto semantico /animato/ pertinente ai due nomi di persona. L'incompatibilità del tratto macrogenerico /animato/ di *Μάρδων, Θάρυβις* nei confronti del tratto /non animato/ di *ἄκμονες* produce un tipo di metafora ben nota fin dall'antichità e detta da Tryph. *trop.* p. 192, 18 *ἀπὸ ἀψύχων ἐπὶ ἔμψυχα*⁵⁵.

Secondo la tassonomia di Genette⁵⁶ questo esempio è classificabile come una «assimilation non motivée», in quanto è presente il «comparé», il «comparant» ma non il «motif» della comparazione. Il *medium comparationis* in questo

⁵⁵Si veda LAUSBERG (1990, p. 287).

⁵⁶In BREMOND (1994, p. 243).

caso si ricava dalle letture del testo: lo scolio⁵⁷ **M** dice ἀκίνητοι ὑπὸ λόγῃς, ὡς ἄκμων ὑπὸ σφυρῶν e lo scolio di Triclinio dice καρτερικοί, ἐκ μεταφορᾶς τοῦ ἄκμονος.

Dall'analisi dei tratti all'impiego di ontologie

Si è visto che il sema o tratto semantico ha natura distintiva: operativamente, ciò significa che in uno studio di semantica vengono identificati tanti tratti quanti sono necessari e sufficienti a differenziare i termini appartenenti a un medesimo campo: l'appartenenza al campo stesso è determinata dalla condivisione da parte di tutti i suoi membri di almeno un tratto microgenerico.

Martínez Hernández ad esempio, nel suo saggio di semantica strutturale-funzionale applicata al «campo léxico de los sustantivos de “dolor” en Sófocles», prende in considerazione i seguenti sostantivi impiegati nelle tragedie del suo autore esaminato: ἄλγος, ἄλγημα, ἄλγησις, ἀλγηδών; πάθος, πάθη, πάθημα, πένθος; μόγος, μόχθος, μόχθημα; ἄχθος, ἄχος; λύπη, λύπημα; ὀδύνη, ὠδύς; πῆμα, πημονή; ἀνία; δύη; κάματος; πόνος. Per condurre l'analisi, egli isola i seguenti tratti distintivi, necessari e sufficienti a opporre ciascun termine a tutti gli altri della collezione: /fisico/, /moral/, /general/, che si riferiscono al dolore del corpo, dell'anima o di entrambi; /por herida o enfermedad/, /por ausencia o muerte de ser querido/, /por acontecimiento externo/, /por insulto u ofensa/, che si riferiscono all'origine del dolore; /implicando compasión/, /implicando arrepentimiento/, /con cólera/, /con odio/, che indicano il modo in cui il dolore si manifesta; /profundo/, che ne indica l'intensità e infine /del parto/, che riguarda un particolare tipo di dolore. In Tab. 6.1 si può vedere una piccola porzione della tavola di contingenza che riassume i risultati ottenuti dallo studioso⁵⁸. Come si può vedere, i tratti possono essere raggruppati e strutturati in base alla funzione che essi esprimono, come ad esempio la causa o il modo. Inoltre la presenza di un tratto può implicare la presenza o, al contrario, l'assenza di un altro tratto: ad esempio

⁵⁷PÉREZ MARTEL (2000) espone il metodo per estrarre e classificare in modo sistematico le informazioni di natura semantica presenti negli scoli e nelle glosse alle tragedie di Eschilo.

⁵⁸La tavola di contingenza completa si può consultare in MARTÍNEZ HERNÁNDEZ (1997, pp. 173-4).

6. Modelli e risorse lessico-semantici

	...	fisico	moral	general	por he- rida o enferme- dad	por ausencia o muerte de ser querido	por aconte- cimiento	...	implicando arrepenti- miento	...
...
πάθος	...	+	+	+	+		
πάθη	...			+			
πάθημα	...	+	+		+	+	+
πένθος	...		+			+	+	...	+	...
...

Tabella 6.1: Analisi componenziale del lessico del dolore in Sofocle

il tratto /por ausencia o muerte de ser querido/ implica il tratto /morale/, mentre, in un contesto dato, i tratti /fisico/, /moral/ e /general/ sono mutualmente esclusivi⁵⁹.

Si è visto all’inizio del capitolo che per il greco antico gli studi di semantica hanno approfondito la conoscenza soltanto di un numero ristretto di campi lessicali⁶⁰: ciò comporta che non vi sia un’opera di riferimento da utilizzare come guida per attribuire a ciascuna entrata del lessico i tratti semantici di pertinenza e potere in questo modo identificare in modo esaustivo tutti i campi semantici messi in gioco sull’asse sintagmatico in un testo determinato.

In assenza di opere sistematiche da cui partire, si è preferito rinunciare, nel nostro lavoro, all’identificazione dei tratti distintivi che ci avrebbero permesso da un lato di determinare i campi lessicali (campo del dolore, campo del timore, etc.), dall’altro di differenziare i termini appartenenti al medesimo campo: gli studi di Martínez Hernández dimostrano come l’articolazione già solo di un singolo campo lessicale richieda una notevole competenza specifica.

Per scegliere quali tratti semantici assegnare a ciascuna occorrenza lessicale del nostro testo si è ricorso quindi non tanto alla nozione di *sema*, quanto piuttosto alla nozione di *noema*. È bene ricordare che il noema, nell’accezione vista sopra,

⁵⁹Non sono mutualmente esclusivi in Tab. 6.1 perché questa tavola di contingenza non distingue le singole occorrenze dei lessemi ma rappresenta invece la somma logica dei tratti manifestati in ciascuna occorrenza

⁶⁰Nonostante Adrados e il suo gruppo, come testimoniano ADRADOS ET AL. (1977) e ADRADOS (2000), stiano curando la redazione del dizionario Griego-Español avendo presente i principi della semantica strutturale, tuttavia non è stato messo in opera alcun inventario dei tratti semantici distintivi.

è un universale di metodo: mentre il sema è un tratto distintivo dipendente dalla specifica forma del contenuto di ciascuna lingua, il noema, al contrario, è stabilito in modo indipendente dalle lingue alle quali si applica e ha lo scopo di organizzare in categorie la sostanza del contenuto a prescindere dalla forma linguistica.

Nell'ambito dell'Intelligenza Artificiale di scuola americana, che si appoggia al paradigma cognitivista, la formalizzazione di una rete concettuale per mezzo di primitive semantiche stabilite a priori (aventi quindi lo statuto di noemi), di costruzione di rapporti gerarchici fra i nodi e di attribuzione di relazioni pre-determinate assume il nome di *ontologia*. Un'ontologia si deve intendere quindi come una descrizione formale di una rete concettuale motivatamente stabilita da una istituzione determinata. Per ora ci interessa soltanto dire che un'ontologia è una tassonomia con finalità pratiche, non speculative. Esistono così ontologie di settore, ad esempio in ambito medico⁶¹, biblio-sitografico⁶², etc. Per i nostri fini ci si è rivolti quindi allo studio di ontologie create in ambito lessicografico.

La nozione di categoria nella prospettiva cognitivista

Nella prospettiva cognitivista⁶³ la nozione di categoria è definita tramite una funzione di appartenenza che ammette infiniti valori fra zero (non appartenenza alla categoria) e uno (piena appartenenza alla categoria). Trasportando in linguistica la teoria degli insiemi sfumati di Zadeh e i risultati degli studi sulla prototipicità della Rosch, Lakoff intende la categoria come un insieme sfumato dotato di un nucleo, costituito da elementi prototipici che possiedono la maggior parte dei tratti caratterizzanti la categoria, e da una zona periferica, dove si collocano gli elementi meno prototipici. Nella sua analisi esemplificativa sul concetto di madre, Lakoff identifica un certo numero di tratti il cui pieno possesso corrisponde al prototipo di madre. La progressiva mancanza di questi tratti stabilisce una distanza concettuale fra il prototipo e l'elemento da comparare: ad esempio la

⁶¹Si veda ad esempio il sito dell'*Institute for Formal Ontology and Medical Information Science - Saarland University*: <http://ifomis.org>

⁶²Si faccia riferimento soprattutto al sito della *Dublin Core Metadata Initiative*: <http://dublincore.org>

⁶³Si veda LAKOFF (1987)

6. Modelli e risorse lessico-semantiche

madre adottiva è, secondo Lakoff, meno prototipica rispetto alla madre che ha generato i propri figli. È ben inteso che la prototipicità deve essere stabilita sperimentalmente o tramite test di psicologia cognitiva oppure tramite l'analisi di *corpora* testuali volta a valutare la cooccorrenza del termine posto sotto esame e di termini che esprimono i tratti associati.

Secondo Lakoff, che parte da una prospettiva semasiologica, la polisemia dei termini è data dalla grande capacità della mente umana di ragionare tramite la somiglianza fra concetti, per metafore⁶⁴ che consistono nel combinare i tratti di categorie diverse e di ignorare parte dei tratti della propria categoria. È grazie a questo procedimento che, tornando al suo esempio, si chiama “madre” tanto la madre naturale quanto la madre adottiva, ma anche la “scena madre” o la “città madre”. Si noti che i tratti che permettono di determinare le categorie, in questa prospettiva, sono proprietà del referente, hanno quindi natura extralinguistica. Così, “madre” si dice in molti sensi, ma le diverse accezioni sono tenute unite da una tessitura di tratti condivisi da più membri, anche nel caso in cui nessun tratto sia condiviso da tutti i membri. Si crea così una rete concettuale che permette di passare da un nodo all'altro e quindi da un concetto a concetti gradatamente sempre più dissimili.

In questa prospettiva cognitivista si sono sviluppati gli strumenti di analisi semantica e lessicografica che andiamo a illustrare.

6.2 WordNet(s)

È chiaro che la redazione di dizionari e ancor più di repertori di sinonimi e contrari prevede l'organizzazione delle voci in base a una ontologia implicita. Tuttavia si può riconoscere a WordNet⁶⁵ il merito di essere il primo progetto informatico

⁶⁴Si veda LAKOFF (1980)

⁶⁵FELLBAUM (1998b) illustra l'architettura di WordNet e i dettagli implementativi. Il progetto è stato sviluppato presso il *Princeton University's Cognitive Science Laboratory* sotto la direzione di G. Miller. Il sito di riferimento del progetto è: <https://wordnet.princeton.edu>, da cui è possibile scaricare il software, trovare i riferimenti a progetti correlati, consultare la documentazione e gli aggiornamenti bibliografici.

con finalità lessicografiche⁶⁶ ad avere ideato una ontologia esplicita per organizzare un numero di vocaboli⁶⁷ in lingua inglese pari a un dizionario tradizionale di dimensioni medio-grandi. Prima della seconda metà degli anni Ottanta erano state progettate ontologie relative a universi di discorso realizzabili con un numero di vocaboli estremamente ridotto, usati per descrivere un universo cognitivo costruito *ad hoc*: uno degli esempi più noti è il programma SHRDLU ideato da WINOGRAD (1972), dove l'analisi semantica del linguaggio naturale è limitata a un universo di solidi geometrici da individuare e ricollocare. La novità di WordNet quindi, nell'ambito dell'Intelligenza Artificiale, è

the *comprehensiveness hypothesis*: that computational linguistics, if it were ever to process natural languages as people do, would need to have available a store of lexical knowledge as extensive as people have⁶⁸.

L'architettura di WordNet

In WordNet sono presi in considerazione soltanto i vocaboli appartenenti alle categorie grammaticali aperte: nomi, verbi, aggettivi e avverbi. Le relazioni stabilite fra i membri delle differenti categorie variano in base alla natura della categoria stessa: ad esempio i rapporti di meronimia e olonimia sono definiti soltanto per i sostantivi, mentre i rapporti di causalità sono definiti soltanto per i verbi, come si vedrà fra breve.

L'organizzazione in synsets

WordNet assume al contempo una prospettiva semasiologica, che permette di attribuire a ciascuna unità lessicale una pluralità di sensi (polisemia), e una prospettiva onomasiologica, che permette di attribuire a un senso una pluralità di unità lessicali (sinonimia). Questo è possibile tramite l'organizzazione delle unità

⁶⁶È definito infatti dai suoi curatori «an electronic lexical database for the English language», come si può vedere nei riferimenti dati alla nota precedente.

⁶⁷Al momento attuale i vocaboli analizzati sono oltre 144.000, tuttavia il progetto è ancora in espansione.

⁶⁸FELLBAUM (1998b, p. XV).

6. Modelli e risorse lessico-semantiche

lessicali in *synsets*⁶⁹, cioè in liste di vocaboli o unità polirematiche alle quali è possibile attribuire un medesimo senso e una medesima glossa. D'altro canto ciascuna unità lessicale può appartenere a differenti *synsets*, in quanto un vocabolo polisemico possiede sinonimi distinti per ciascun senso preso in esame.

Si prenda come esempio l'aggettivo *luxurious*: in WordNet 2.0 il vocabolo appartiene a due *synsets* differenti. Nel senso di «furnishing gratification of the senses» appartiene al *synset* che ha per membri «epicurean, luxurious, sybaritic, voluptuary, voluptuous»; nel senso invece di «rich and superior in quality» appartiene al *synset* «deluxe, gilded, luxurious, opulent, princely, sumptuous». In greco si può fare un parallelo con il polisemico *ἀβρός*. In Suda *s.v.* abbiamo: *ἀβρός λαμπρός, τρυφερός, ἀπαλός*, informazioni semantiche che si possono ricondurre ad almeno tre differenti *synsets*: «*ἀβρός, λαμπρός, (φαιδρός), ...*» nel senso di “splendido (in senso figurato)”, «*ἀβρός, τρυφερός, (μαλακός), ...*» nel senso di “dedito alle mollezze” e «*ἀβρός, ἀπαλός, (μαλακός), ...*» nel senso di “delicato”.

L'identità di un *synset* in WordNet non è data dalla lista dei suoi membri (infatti due *synsets* con gli stessi membri potrebbero avere sensi differenti) ma da un identificatore univoco costituito da un numero progressivo. Come si è detto, a ciascun *synset* sono associati una glossa, cioè una breve descrizione informale del *synset* espressa in linguaggio naturale, per darne la definizione del senso ed eventualmente qualche esempio d'uso, e una rete di relazioni verso altri *synsets*, come si vedrà in seguito.

Ciascuna categoria grammaticale (nome, verbo, aggettivo e avverbio) è ripartita in sottocategorie semantiche, elencate nella tavola della pagina seguente, che hanno lo scopo pratico di suddividere i *synsets* in *files* distinti, più facilmente leggibili da parte degli operatori coinvolti nel progetto⁷⁰. Per i sostantivi, l'organizzazione in *files* è conforme alla struttura gerarchica dell'ontologia adottata, come si vedrà nei prossimi paragrafi. Anche per i verbi la ripartizione è funzionale all'organizzazione gerarchica dei *synsets*, mentre gli aggettivi sono suddivisi

⁶⁹ *Synset* è l'abbreviazione di *synonym set*.

⁷⁰ FELLBAUM (1998b, p. 28): «Partitioning [...] has one important practical advantage: it reduces the size of the files that lexicographers must work with and makes it possible to assign the writing and editing of different files to different people».

soltanto in due categorie, senza ulteriori suddivisioni e gli avverbi sono infine raccolti tutti in un unico *file*.

Il Sostantivo

Per i sostantivi i nomi dei *files* in cui sono ripartiti i *synsets* rappresentavano, nello stato di progettazione iniziale di WordNet, anche i nodi gerarchicamente più alti della tassonomia alla quale appartiene ciascun *synset*, e per questo erano detti *unique beginners*:

Since the features that characterize a unique beginner are inherited by all of its hyponyms, a unique beginner corresponds roughly to a primitive semantic component in a compositional theory of lexical semantics.⁷¹

Tuttavia i responsabili si sono accorti ben presto che era opportuno raggruppare alcuni di questi presunti *unique beginners*, considerandoli come istanze di concetti più generali.

Così, ad *animal (fauna)*, *person (human being)* e *plant (flora)* è stato assegnato l'iperonimo *organism*; *body* è stato considerato iponimo di *natural_object*, così come *food* di *substance*. A loro volta *natural_object*, *substance* e *artifact* sono diventati iponimi di *object*. Come nodo terminale di questa gerarchia di nomi concreti, che raggruppa *organism* e *object*, è stato assegnato l'iperonimo *entity*. Sono state poi raggruppate sotto l'iperonimo *abstraction* le seguenti sottocategorie: *attribute quantity (amount) relation* (con l'iponimo *communication*) e *time*. Infine, sono state subordinate gerarchicamente a *psychological_feature* le categorie: *cognition (knowledge)*, *feeling (emotion)* e *motivation (motive)*. A esclusione di *process*, subordinato a *phenomenon*, le rimanenti sottocategorie, cioè *activity (act)*, *event (happening)*, *group (grouping)*, *location*, *possession*, *shape* e *state* sono rimaste *unique beginners*.

Questa nuova organizzazione degli *unique beginners* fa emergere con maggiore chiarezza le opposizioni macrogeneriche di /concreto/ vs /astratto/ e /animato/ vs /inanimato/. Per comodità, alla disposizione gerarchica dei nuovi e dei vecchi *unique beginners* daremo il nome di *Top Ontology*.

⁷¹FELLBAUM (1998b, *ibidem*).

6. Modelli e risorse lessico-semantiche

L'organizzazione gerarchica prevista per i sostantivi in WordNet permette di identificare i rapporti di iponimia e iperonimia che ciascun *synset* stabilisce rispetto ad altri *synsets*. Come si può osservare, l'iponimia non è una relazione lessicale fra vocaboli, ma una relazione semantica fra liste di sinonimi (i *synsets* appunto) a cui i vocaboli appartengono.

Un esempio per chiarire. In WordNet 2.0 il vocabolo «horse» appartiene a cinque diversi *synsets* nominali e a un *synset* verbale, del quale qui non ci occupiamo⁷².

I *synsets* di sostantivi corrispondono a cinque differenti sensi del termine, e precisamente:

Sense #1:

{horse, Equus_caballus}: solid-hoofed herbivorous quadruped domesticated since prehistoric times.

Sense #2:

{horse}: a padded gymnastic apparatus on legs.

Sense #3:

{cavalry, horse_cavalry, horse}: troops trained to fight on horseback; “500 horse led the attack”.

Sense #4:

{sawhorse, horse, sawbuck, buck}: a framework for holding wood that is being sawed.

Sense #5:

{knight, horse}: a chessman in the shape of a horse's head; can move two squares horizontally and one vertically (or vice versa).

La gerarchia degli iperonimi relativa al senso #1 è la seguente:

{horse, Equus_caballus} < {equine, equid} < {odd-toed_ungulate, perissodactyl, perissodactyl_mammal} < {ungulate, hoofed_mammal} < {placental, placental_mammal, eutherian, eutherian_mammal} < {mammal} < {vertebrate, craniate} < {chordate} < {animal, animate_being, beast, brute, creature, fauna} < {organism, being} < {living_thing, animate_thing} < {object, physical_object} < {entity}

⁷²Il verbo «to horse» è glossato in WordNet con «provide with a horse or horses»

mentre, tralasciando per brevità gli altri sensi, la gerarchia degli iperonimi del senso #3 è la seguente:

$$\begin{aligned} & \{\text{cavalry, horse_cavalry, horse}\} < \{\text{military_personnel, soldiery, troops}\} \\ & < \{\text{force, personnel}\} < \{\text{organization, organisation}\} < \{\text{social_group}\} < \\ & \{\text{group, grouping}\} \end{aligned}$$

Si è scelto appositamente questo esempio perché permette di fare alcune considerazioni, in primo luogo sulla traducibilità da una lingua all'altra, tema sul quale si tornerà nella sezione dedicata a EuroWordNet, e che a noi interessa perché gli strumenti lessicali utilizzati per il nostro lavoro sono bilingui: Greco-Latino (Italie, Stephanus), Greco-Inglese (*LSJ*), Greco-Spagnolo (Adrados). Se si traduce *horse* in un'altra lingua moderna, ad esempio in Italiano, è evidente che non c'è un vocabolo che si adatti contemporaneamente ai primi quattro sensi visti sopra, i quali corrispondono invece rispettivamente ai vocaboli italiani: *cavallo*, *cavallina*, *cavalleria*, e *cavalletto* (mentre anche in italiano *cavallo* può significare *cavallo degli scacchi*).

Questo problema in WordNet trova una soluzione nella separazione fra unità lessicali (i vocaboli) e unità di significato (i *synsets*). La struttura ontologica del *synset*, espressa dalla gerarchia degli iponimi, vuole avere un contenuto cognitivo, indipendente dal contenuto linguistico: la struttura gerarchica del senso #1 e la struttura del senso #3 fanno riferimento, nella prospettiva cognitivista soggiacente a WordNet, a una realtà extralinguistica: questo fa in modo che il vocabolo italiano *cavallo* possa essere associato correttamente al primo e al quinto *synset*, mentre il vocabolo inglese *horse* sia associato a tutti i cinque *synsets* visti sopra.

Ma il paradigma cognitivista non tiene conto della distinzione hjelmsleviana⁷³ fra materia, forma e sostanza del contenuto e questo fatto, come si vedrà fra poco, assume grande rilievo nell'applicazione della struttura di WordNet al caso specifico del nostro studio. Per Hjelmslev, come per de Saussure, la materia è la massa amorfa del pensiero, un *continuum* extralinguistico. All'interno di una stessa zona della materia (in danese *meningszone*) soltanto la forma linguistica, indipendente dalla materia e avente con essa un rapporto arbitrario, è in grado

⁷³Si veda HJELMSLEV (1993, pp. 46-50).

6. Modelli e risorse lessico-semantiche

di ritagliare diverse sostanze del contenuto: come è noto gli esempi hjelmsleviani riguardano la differente ripartizione del *continuum* delle sfumature del colore verde in gallese e in danese oppure del *continuum* della temporalità nella differente organizzazione dei tempi verbali nelle diverse lingue, etc. Nel paradigma cognitivista invece la concettualizzazione è dichiarata come indipendente dalla lingua che la esprime. Ma fa notare Rastier in *Sémantique et recherches cognitives*:

Mais les «concepts» définis par ces réseaux [*scil.* réseaux sémantiques] sont tout simplement des contenus linguistiques, et on ne doit pas assimiler sens et connaissance sans s'interroger sur leur rapport.

Les rapports entre *sens* et *connaissance* dérivent du postulat que la connaissance est une représentation symbolique du réel⁷⁴.

Rastier qui ha in mente reti semantiche dove ciascun nodo che rappresenta un concetto è etichettato con un singolo vocabolo in una lingua determinata, come ad esempio *horse* per etichettare il nodo che rappresenta il concetto di cavallo. La separazione fra vocaboli e *synsets* in parte supera questa critica, anche se il nodo associato al *synset* è comunque descritto tramite la lista dei vocaboli e dalla glossa a esso associati, espressi per forza di cose in una lingua data. Ma il punto che qui più ci interessa è la natura dei legami fra i *synsets*:

Certains admettent que l'ensemble des liens définit le sens des concepts (cf. Sabah, 1984, p. 481⁷⁵). Cependant, les «liens sémantiques» en question ne sont pas dépendants d'une langue déterminée ; par exemple, le lien *is-a* (*sorte de*) exprime la relation entre un ensemble et ses éléments (qui n'est propre à aucune langue en particulier).

Dunque, se si accetta questa posizione, i rapporti di iponimia/iperonimia (e, si può aggiungere, di meronimia/olonimia) sono validi indipendentemente dall'espressione linguistica:

Les liens relèvent donc du niveau conceptuel et leur existence est purement métalinguistique. C'est pourquoi d'ailleurs ils paraissent se traduire ou se transposer parfaitement d'une langue à l'autre, à la différence des étiquettes des nœuds⁷⁶.

⁷⁴RASTIER (2001, p. 36).

⁷⁵Si fa riferimento a Sabah, Différentes notions de synonymies liées à la compréhension automatique du langage, *Les Modes de Raisonnement*, Actes du second colloque de l'A.R.C., pp. 477-504.

⁷⁶RASTIER (2001, p. 128).

Questa ultima affermazione di grande portata nasconde tuttavia un'insidia: la traduzione o la trasposizione di una concettualizzazione da una lingua all'altra è efficace soltanto se la semantica di enciclopedia è condivisa, cioè se le lingue in causa sono espressione di civiltà che hanno un medesimo orizzonte scientifico e culturale.

Nel campo della lessicografia e della semantica computazionale la tassonomia di WordNet costituisce comunque un punto di riferimento quasi imprescindibile. Si è ritenuto quindi opportuno fare questa lunga digressione teorica per dare una giustificazione ai criteri generali che si sono seguiti nell'adottare o nel rifiutare, caso per caso, le soluzioni proposte dai responsabili di WordNet.

Nel testo dei *Persiani* ἵππος occorre quattro volte, in accordo con *Italie* due volte nel senso di *equus* al v. 18 τοὶ μὲν ἐφ' ἵππων, τοὶ δ' ἐπὶ ναῶν, πέλζοί τε βάρην e al v. 32 ἵππων τ' ἐλατήρ Σοσθάνης e due volte nel senso di *equitatus*: al v. 302 Ἀρτεμβάρης δὲ μυρίας ἵππου βραβεύς e al v. 315 ἵππου μελαίνης ἡγεμῶν τρισμυρίας⁷⁷.

Per un caso fortunato, che è anche il motivo della scelta di questo esempio, la polisemia del greco ἵππος ha due punti di intersezione con la polisemia dell'inglese *horse*: le occorrenze dei vv. 18 e 32 trovano corrispondenza con il senso #1 di *horse* e le occorrenze dei vv. 302 e 315 trovano corrispondenza con il senso #3. Tuttavia, come accade sistematicamente in WordNet, l'ordinamento gerarchico per regno, phylum, classe, ordine, famiglia, genere e specie (con eventuali ulteriori suddivisioni, come sottoclasse o sottordine) degli animali riproduce la tassonomia linneiana, non pertinente all'orizzonte scientifico-culturale dell'Atene del V secolo.

Ai fini del nostro lavoro, ci si è discostati dalla tassonomia di WordNet in tutti i casi in cui l'organizzazione gerarchica dei *synsets* non è applicabile ai domini di conoscenza o di organizzazione sociale del V secolo. Si è detto sopra, parlando della differenza fra dimensioni, domini e tassemi, che le dimensioni (ad esempio concreto/astratto, animato/inanimato) subiscono variazioni nell'ordine dei millenni mentre i domini (zoologico, botanico, militare, marinaro, etc.) si

⁷⁷A proposito di μελαίνης GROENEBOOM (1930, *ad loc.*), osserva: «het epitheton slaat natuurlijk op de ruiters, want dat negers geen baard hebben wisten ook de Grieken».

6. Modelli e risorse lessico-semantiche

evolvono con l'evolversi della scienza e della tecnica a essi relative, e quindi hanno variazioni nell'ordine dei secoli se non ancora più brevi.

Operativamente, si è scelto di accogliere la ripartizione degli *unique beginners* proposta da WordNet (che, come si è visto, ha forti analogie con la nozione di dimensione semantica di cui parla Rastier), mentre si è proceduto a una notevole riduzione degli iperonimi intermedi fra il *synset* analizzato e i nodi della *Top Ontology*.

Si ricorda che nel nostro lavoro il ricorso all'organizzazione del lessico all'interno di una tassonomia è funzionale allo studio di cooccorrenze o di opposizioni di tratti pertinenti. Rispetto alla tassonomia di WordNet si sono registrati quindi soltanto gli iperonimi pertinenti al testo. Questo avviene in due casi: o quando l'iperonimo e uno o più iponimi sono espressi nel testo (ad esempio: *τοκέυς* e *μήτηρ, πατήρ*) oppure nel caso in cui l'iperonimo, anche se non espresso nel testo, ha il carattere di tasema nei confronti di due o più coiponimi espressi nel testo. Due esempi: benché nel testo non sia impiegato alcun termine con il significato specifico di "relative", cioè di «person related by blood or marriage», tuttavia i vv. 63-4:

*τοκέες τ' ἄλοχοί θ' ἡμερολεγδὸν
τείνοντα χρόνον τρομέονται.*

ci permette di postulare un nodo della rete semantica a cui tanto *τοκέες* quanto *ἄλοχοι* sono subordinati. Tuttavia, la descrizione in inglese che si è scelta per questo nodo non è "relative" ma "relative or lover" (comprendendo legami di sangue, matrimoniali e affettivi), in quanto per il nodo che unisce i due *synsets* {*ἀνὴρ, εὐνητήρ, πόσις*} e {*ἄκοιτις, ἄλοχος, γυνή, εὐνήτεια*} si è scelta la descrizione inglese di "significant other", che si estende tanto a "partner" quanto a "lover"⁷⁸.

⁷⁸In Soph. *Tr.* 550s.

*Ταῦτ' οὖν φοβοῦμαι μὴ πόσις μὲν Ἡρακλῆς
ἐμὸς καλῆται, τῆς νεωτέρας δ' ἀνὴρ.*

πόσις e *ἀνὴρ* sono esplicitamente distinti. Dunque, la ragione per cui si è scelta la descrizione di "significant other" è il fatto che nella nostra tragedia non vi sono né contesti dove sia tematizzata la distinzione /legame legittimo/ *vs* /legame illegittimo/ né contesti che permettano di stabilire

Il secondo esempio riguarda *Pers.* 205–7, che oppone *ἀετός* a *κίρκος* come simboli rispettivamente del Persiano e del Greco. In questo caso il tassema che li accomuna è descritto da “raptorial bird” e non semplicemente da “bird” perché la distinzione è pertinente al contesto: la particolarità del *τέρας* raccontato dalla regina consiste nel fatto che un rapace che usualmente attacca colombe⁷⁹ in questo caso abbia attaccato e vinto un altro rapace.

Per i motivi appena visti, dovrebbe essere chiaro che nel nostro lavoro gli iperonimi sono rappresentati da *synsets* vuoti in tutti i casi in cui non vi è, nel testo analizzato, un vocabolo o un gruppo di vocaboli che possa essere associato al nodo che riunisce i coiponimi. In tal caso, al nodo è associata soltanto una descrizione in lingua inglese. Un’apposita funzione di ricerca definita in WordNet permette di individuare, a partire da un vocabolo dato, i coiponimi dei suoi *synsets* di appartenenza.

Il secondo tipo di relazione gerarchica codificata in WordNet è il rapporto di meronimia/olonimia. La relazione che intercorre fra la parte e il tutto può essere di diversi tipi: componente/oggetto, membro/gruppo, materia/manufatto, porzione/massa, fase/attività, luogo/area, etc., tuttavia in WordNet sono codificate soltanto le prime tre relazioni, con le seguenti definizioni⁸⁰:

- x is a (component) part of y
- x is a member of y
- x is the stuff that y is made from

Altri tipi di relazione meno prototipici sono quindi ricondotti alla prima: *x is a part of y*. In modo simile a WordNet, dove la maggior parte dei meronimi/olonimi sono concentrati, nella *Top Ontology*, all’interno delle categorie **noun.body**, **noun.artifact** e **noun.quantity**⁸¹, anche nel nostro lavoro la maggior parte dei

in modo inequivocabile che tutti i termini più generici (*άνήρ, ἄκοιτις, γυνή*) sono impiegati nel senso restrittivo di /legame legittimo/.

⁷⁹Il motivo tradizionale si trova ad esempio in *PV* 857 *κίρκοι πελειῶν οὐ μακρὰν λελειμμένοι* e in *Supp.* 223–4 *ὡς πελειάδων / ἴζεσθε κίρκων τῶν ὀμοπτέρων φόβῳ*, come figura in entrambi i passi dei cugini egizi contro le Danaidi.

⁸⁰Si veda G.A. Miller in FELLBAUM (1998b, pp. 37–41).

⁸¹Per quanto riguarda la categoria **noun.quantity**, WordNet contempla i rapporti di olonimia/meronimia fra le unità di misura e le sottounità.

6. Modelli e risorse lessico-semantiche

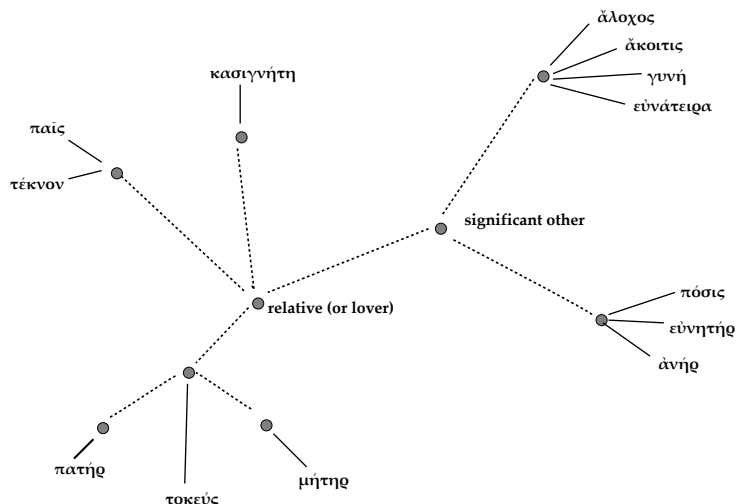


Figura 6.1: Organizzazione dei *synsets* e relazioni di iponimia/iperonimia.

meronimi/olonimi appartengono alle categorie *noun.body* (parti del corpo umano e animale) e *noun.artifact* (parti della nave e delle armi). Al contrario delle relazioni di iponimia/iperonimia e di meronimia/olonimia, che in WordNet sono istituite fra significati e non fra vocaboli, il rapporto di antonimia invece è stabilito fra singoli vocaboli che appartengono a un *synset*. La ragione si trova, ancora una volta, nell'impostazione cognitivista⁸²:

The strongest psycholinguistic indication that two words are antonyms is that each is given on a word association test as the most common response to the other.

Un motivo sperimentale dunque, che nel nostro lavoro è stato sostituito dalla pertinenza in contesto, da ragioni etimologiche e dall'opposizione all'interno del medesimo tassema, conclusione questa a cui perviene, partendo da premesse diverse, lo stesso Miller⁸³:

⁸²Si veda FELLBAUM (1998b, *ibid.*).

⁸³FELLBAUM (1998b, *ibid.*).

Perhaps the most interesting observation about antonymous nouns is that noun antonyms nearly always have the same hypernym, often the same immediate hypernym.

Mettendo insieme tutte le considerazioni fatte, si prenda come esempio la coppia di *synsets* vista sopra, che ha come iperonimo il nodo semantico etichettato come “significant other”. Poiché il rapporto di antonimia è stabilito fra vocaboli e non fra nodi concettuali, le coppie di antonimi che si vengono a stabilire, seguendo i criteri appena visti, sono: ἀνήρ/γυνή, in quanto coppia tradizionale di opposti, εὐνητήρ/εὐνάτειρα, per il loro legame etimologico, πόσις/ἄλοχος, grazie a luoghi paralleli che stabiliscono delle simmetrie: e.g. Hom. *E* 414-5 ποθέουσα πόσιν τὸν ἄριστον / ἰφθίμη ἄλοχος Διομήδεος ἵπποδάμοιο oppure Eur. *Alc.* 165-6 καὶ τῷ μὲν φίλην / σύζευξον ἄλοχον, τῇ δὲ γενναῖον πόσιν. Si tornerà comunque sui dettagli di queste opposizioni nei capitoli successivi.

In WordNet i nomi astratti che indicano degli attributi, come *size*, *color*, etc. sono posti in relazione con gli aggettivi che rappresentano i valori di tali attributi, come ad esempio *large*, *big*, *small* e *little* per il primo e *red*, *green*, *blue*, etc. per il secondo. Nel nostro lavoro ci si è attenuti per lo più a un criterio di legame etimologico, mettendo in relazione termini astratti come μέγεθος, κάλλος, τάχος rispettivamente con gli aggettivi μέγας, καλός, ταχύς e i loro rispettivi eventuali antonimi.

Il Verbo

Anche i verbi, come i sostantivi, hanno un’organizzazione gerarchica che permette di individuare rapporti di iponimia e iperonimia: al vertice della gerarchia ci stanno i nodi della *Top Ontology* (verb.cognition, verb.communication, verb.motion, etc.), che ricordano abbastanza da vicino la suddivisione tradizionale in *verba putandi*, *verba dicendi*, *verba movendi* etc., anche se in questo caso la ragione del raggruppamento non è legata alla similarità nella costruzione sintattica, ma a motivi prettamente semantici.

Il rapporto di iponimia è stabilito in base alla maggiore specificità di un significato verbale rispetto a un altro, in quanto l’iponimo possiede tutti i tratti

6. Modelli e risorse lessico-semantiche

dell'iperonimo più alcuni suoi propri, come ad esempio l'indicazione dell'intensità, della finalità, della modalità, etc. Vediamo due esempi. Nel nostro testo il verbo polisemico *καλέω* può essere fatto rientrare in tre differenti *synsets*, a seconda del contesto: al v. 2 il coro di anziani dice di sé stesso: *πιστὰ καλεῖται*, mentre al v. 242 il messaggero dice dei Greci: *οὔτινος δοῦλοι κέκληνται φωτὸς*. Accogliamo qui le osservazioni di Groeneboom al v. 2 e al v. 242, riprese nell'affermazione di MATINO (1998, p. 133):

Per il profondo legame tra l'essere chiamato e l'esistere, nell'epica e in Eschilo *καλέω* corrisponde ad *εἰμί*.

Di conseguenza il primo *synset* comprende {*εἰμί, καλέω, πέλω*}. Al v. 175 invece il coro si rivolge alla regina, che ha appena chiesto consiglio, e dice: *εὐμενείς γὰρ ὄντας ἡμᾶς τῶνδε συμβούλους καλεῖς*. Questa volta l'uso del verbo ha il pieno significato di *vocare*, e rientra in un *synset* con un solo componente: {*καλέω*}. Infine, ai vv. 687-8 Dario parla del modo in cui il coro lo sta evocando dall'oltretomba: *οἰκτρῶς καλεῖσθέ μ(ε)*. Questa azione è il compimento dell'ordine dato al coro dalla regina ai vv. 620-1: *ῥμνους ἐπευφημεῖτε, τόν τε δαίμονα / Δαρείον ἀνακαλεῖσθε*. Il terzo *synset* comprende dunque {*ἀνακαλέω, καλέω*}. Fra il valore più specifico di *evocare* e il valore più generale di *vocare* si stabilisce dunque il rapporto di iponimia. Questo caso è simile a quello visto sopra: là *ἀνῆρ* nel senso di *maritus* è iponimo di *ἀνῆρ* nel senso di *homo*, qui *καλέω* nel senso di *evocare* è iponimo di *καλέω* nel senso di *vocare*. Il rapporto di iponimia fra due *synsets* che contengono il medesimo vocabolo rende esplicita la presenza di una catacresi sineddohica del genere e della specie.

Come secondo, più sbrigativo esempio si può osservare il rapporto di iponimia fra l'accezione generica di *ὄράω* e le accezioni specifiche dei verbi *προσοράω, εἰσοράω, δέρκομαι, καθοράω, βλέπω, λεύσσω*⁸⁴.

Come per i sostantivi, anche per i verbi è possibile, a partire da un vocabolo, individuare i coiponimi relativi ai suoi *synsets* di appartenenza.

⁸⁴Si fa riferimento alle note considerazioni sull'argomento, relativamente all'epica omerica, da parte di SNELL (1955).

Per i verbi WordNet non prevede il rapporto di meronimia/olonomia, bensì la relazione di *entailment*, che si verifica quando la nozione espressa da un verbo implica la nozione espressa da un altro verbo, che la include: stanno in questo rapporto ad esempio i verbi incoativi o egressivi rispetto ai verbi che esprimono l'azione nella sua interezza. Altro rapporto di implicazione logica è la relazione di causa/effetto: un caso⁸⁵ è *κτείνω* rispetto a *θνήσκω*.

Come per il sostantivo, anche per il verbo il rapporto di antonimia è istituito fra vocaboli e non fra nodi della rete concettuale. Fra gli altri, sono considerati antonimi i verbi che derivano da sostantivi o da aggettivi che sono fra di loro antonimi, come *πίμπλημι* e *κενώνω*, oppure verbi dove vi sia uno scambio di ruolo tematico, come *διδάσκω* e *μανθάνω*.

L'aggettivo e l'avverbio

Gli aggettivi non sono organizzati in una struttura gerarchica come quella di sostantivi e verbi, ma sono suddivisi soltanto in due grandi categorie: relazionali e non relazionali, fra i quali quelli detti descrittivi. Sono considerati relazionali (*pertainyms*) gli aggettivi derivati da un sostantivo *x* che possono essere trasformati nell'espressione: "relativo, pertinente a *x*", oppure in un genitivo di specificazione. È il caso di *βασίλειος* (*Pers.* 8, 66, etc.) rispetto a *βασιλεύς* (5, 24, etc.), di *δεσπόσυνος* (587) rispetto a *δεσπότης* (169, 666, 1049) di *τοξικός* (460) rispetto a *τόξον* (147, 278), etc.

I «descriptive adjectives»⁸⁶ sono considerati valori di un attributo: così, come si è visto sopra, *μέγας* è un valore dell'attributo *μέγεθος*. Quando non sia presente nel testo alcun vocabolo da associare al nodo della rete semantica che rappresenta l'attributo corretto, il nodo viene descritto in lingua inglese: ad esempio *ἀγαθός* e *κακός*, in senso non morale, sono valori dell'attributo *quality*.

Il rapporto privilegiato fra aggettivi è la relazione di antinomia, che permette di strutturarli in coppie di contrari e che, per ciascun polo della coppia, permette

⁸⁵Si vedano tuttavia i problemi sollevati dall'analisi di «kill» come «cause to die» in FODOR (1970).

⁸⁶Si veda FELLBAUM (1998b, pp. 47–66).

6. Modelli e risorse lessico-semantiche

di organizzare una rosa di aggettivi di significato simile ma privi di un antonimo diretto.

Thus, *ponderous* is similar to *heavy* and *heavy* is the antonym of *light*, so a conceptual opposition of *ponderous/light* is mediated by *heavy*. Gross, Fischer, and Miller distinguish direct antonyms like *heavy/light*, which are conceptual opposites that are also lexically paired. Under this formulation, all descriptive adjectives have antonyms; those lacking direct antonyms have indirect antonyms (i.e., are similar in meaning to adjectives that have direct antonyms)⁸⁷.

Nella prospettiva cognitivista, è possibile verificare l'antonimia diretta sperimentalmente, in quanto la quasi totalità dei parlanti nativi sottoposta a *test* è portata ad associare per antonimia *heavy* a *light* e non ad *airy*, oppure *light* a *heavy*, non a *ponderous*. Benché l'associazione sia lessicale, tuttavia un aggettivo polisemico può avere più di un antonimo diretto, in base al senso: *light* infatti si oppone a *heavy* come valore dell'attributo *weight*, ma a *dark* come valore dell'attributo *lightness*.

Il fatto che all'antonimia sia riconosciuta una natura prevalentemente lessicale e non concettuale rende l'identificazione dei contrari possibile anche in assenza di parlanti nativi da sottoporre a test che richiedano la loro *competence* linguistica: è possibile infatti costituire le coppie di contrari a partire dalle *performances*, in particolare da basi di dati testuali. È stato dimostrato infatti da JUSTESON E KATZ (1991) che le stesse coppie di antonimi formate dai parlanti nativi durante i test sono rilevate tramite procedure di statistica testuale. Si vede infatti che la frequenza della cooccorrenza di antonimi in un *corpus* sincronico di testi è maggiore ai valori attesi se la cooccorrenza fosse casuale. Questo fornisce quindi un criterio oggettivo per procedere all'associazione degli antonimi: è l'analisi dei *loci paralleli* in cui gli antonimi sono congiunti o disgiunti che giustifica la loro costituzione in coppie lessicali di opposti.

La tradizione filosofica del resto ci ha abituati a riconoscere coppie di opposti lessicalmente stabili. Per fare un esempio fra i molti, basti prendere Arist. *Met.* 1089b:

⁸⁷FELLBAUM (1998b, 50).

καίτοι χρώνται καὶ λέγουσι μέγα μικρόν, πολὺ ὀλίγον, ἐξ ὧν οἱ ἀριθμοί, μακρὸν βραχύ, ἐξ ὧν τὸ μῆκος, πλατὺ στενόν, ἐξ ὧν τὸ ἐπίπεδον, βαθὺ ταπεινόν, ἐξ ὧν οἱ ὄγκοι

Tuttavia l'applicazione a testi poetici invita alla massima prudenza: ragioni metriche e stilistiche spingono alla cooccorrenza nel testo di coppie di opposti lessicalmente differenti ma identiche dal punto di vista del significato. Così, ad esempio, non si può dire, usando i termini di WordNet, che *παῦρος* è un antonimo indiretto di *πολύς* per mezzo della sua similarità di significato con *ὀλίγος*. Almeno per quanto riguarda la Tragedia è più prudente dire che *παῦρος* e *ὀλίγος* sono concorrenti nel loro rapporto di antonimia nei confronti di *πολύς*.

Vediamo l'uso eschileo dell'uno e dell'altro aggettivo, integrando con alcune considerazioni sugli altri tragici, per dimostrare quanto appena detto. *ὀλίγος* occorre in *Pers.* 330 e *Sept.* 762, *παῦρος* occorre invece in *Pers.* 800, *Ag.* 832, *Fr.* 25e e *Fr.* 99 Radt, ma a noi interessano soltanto i luoghi in cui uno dei due aggettivi è in opposizione a *πολύς*.

L'occorrenza di *ὀλίγος* in *Pers.* 330 contrappone la moltitudine dei mali presenti al piccolo numero di quelli annunciati dal messo: *πολλῶν παρόντων δ' ὀλίγ' ἀπαγγέλλω κακά*. Così almeno nell'edizione di West, ma altri editori, fra cui Wecklein (e Murray) omettono, in accordo con i codici **MD**, il δ'. Senza la particella anche *παῦρ'* sarebbe compatibile con il contesto metrico e semantico e infatti Wecklein osserva, nella sua *Appendix*: «fort(asse) *παῦρ'*»: proposta interessante, perché *difficilior* rispetto a *ὀλίγ'*, per il fatto che troverebbe un parallelo in *Pers.* 800 anche nell'allitterazione di π, e per il fatto che troverebbe altri paralleli anche nel *Fr.* 99 Radt *τὰ πολλὰ κεῖνα διὰ παύρων λέγω*. oppure nel verso sofocleo di *Electr.* 688 *Χῶπως μὲν ἐν πολλοῖσι παῦρά σοι λέγω*. Sofocle del resto predilige *παυρός* anche in *Aj.* 1022: *πολλοὶ μὲν ἐχθροί, παῦρα δ' ὠφέλησιμα*.

Ma l'idea di Wecklein purtroppo non è sostenuta da varianti testuali e quindi è bene accogliere il tradito *ὀλίγ'*, che d'altro canto dà anche maggiore vivacità al ritmo del verso.

In *Pers.* 800-2 alle richieste del coro se tutto l'esercito abbia attraversato l'Ellesponto, Dario risponde:

6. Modelli e risorse lessico-semantiche

παῦροί γε πολλῶν, εἴ τι πιστεῦσαι θεῶν
χρῆ θεσφάτοισιν, εἰς τὰ νῦν πεπραγμένα
βλέψαντα συμβαίνει γὰρ οὐ τὰ μέν, τὰ δ' οὔ.

In questo caso è *παῦρος* invece di *ὀλίγος* a essere contrapposto a *πολύς*, e anche in questo caso non vi sono necessità metriche nella scelta di uno dei due aggettivi concorrenti. Per *ὀλίγος* in prima sede da opporre a *πολύς* propende invece Euripide, in *Med.* 120 *ὀλίγ' ἀρχόμενοι, πολλὰ κρατοῦντες*, nel frammento papiraceo Fr. 18 Austin *ὀλίγοι γὰρ ἐσθλοὶ κρείσσονες πολλῶν κακῶν* e nel Fr. 46 Austin *ὀλίγους ἐπαινῶ μᾶλλον ἢ πολλοὺς κακοῦς*, ma in *Heracl.* 327 preferisce l'altro aggettivo: *παύρων μετ' ἄλλων ἕνα γὰρ ἐν πολλοῖς ἴσως*.

Riassumendo il quadro, si può dire che è preferibile considerare, in Tragedia, entrambi i sinonimi *παῦρος* e *ὀλίγος* come antonimi diretti di *πολύς*.

Per finire, un breve cenno all'organizzazione degli avverbi, che in WordNet, come gli aggettivi, non sono organizzati in una tassonomia. L'unico tipo di relazione è stabilito a livello di vocabolo e non di *synset* con l'aggettivo da cui l'avverbio deriva.

EuroWordNet

EuroWordNet⁸⁸ è nata come estensione indipendente⁸⁹ di WordNet 1.5 per coprire le esigenze di una rete semantica in un contesto multilingue. Oltre alla creazione di un dizionario elettronico multilingue, lo scopo principale del progetto era quello di fornire uno strumento di sostegno alla traduzione semiautomatica di testi scritti nelle differenti lingue europee.

L'architettura di EuroWordNet si basa sull'idea di utilizzare i *synsets* e le glosse di WordNet, con opportuni aggiustamenti e integrazioni, per costituire una base concettuale indipendente dalle diverse lingue coinvolte nel progetto. Poiché,

⁸⁸EuroWordNet è un progetto finanziato dalla Comunità Europea dal 1996 al 1999. Documentazione e materiali si trovano al seguente indirizzo: <http://www.illc.uva.nl/EuroWordNet>.

⁸⁹Tutti i materiali di WordNet sono distribuiti come *open source*. È possibile modificare ed estendere come progetti indipendenti tanto la base di dati quanto il software di interrogazione.

come si è visto sopra, i rapporti di iponimia/iperonimia fra i nodi di una rete concettuale sono condizionati dalla lingua che li esprime, in questa base comune «we get a much flatter hierarchy in which particular properties such as *natural*, *artificial* and *functionality* cannot be derived from the hyponymy relations»⁹⁰. Lo scopo è di avere una collezione di nodi concettuali che costituisca un *Inter-Lingual-Index* al quale sono associati, mediante una funzione di equivalenza, i *synsets* di ciascuna base di dati monolingue: in questo modo i *synsets* specifici di ciascuna lingua che siano collegati a un medesimo nodo concettuale dell'*Inter-Lingual-Index* sono considerati possibili traduzioni degli uni rispetto agli altri.

Il rapporto di iponimia/iperonimia, a livello microgenerico, è stabilito nelle reti semantiche riguardanti ciascuna lingua specifica, in quanto si è deciso che la mancanza di un iperonimo in una lingua non può essere sostituito da una perifrasi soltanto perché in un'altra lingua l'iperonimo esiste: ad esempio in lingua inglese i nodi concettuali che indicano diversi tipi di contenitori sono gerarchicamente subordinati al nodo *container*. In Olandese tuttavia non esiste alcun termine con questo significato: il concetto di contenitore potrebbe essere espresso soltanto con una perifrasi. In EuroWordNet quindi la base di dati inglese e la base di dati olandese si comportano in maniera diversa nello stabilire quale sia l'iperonimo dei termini che indicano i contenitori.

A livello macrogenerico invece l'organizzazione gerarchica viene fatta su alcuni nodi particolari dell'*Inter-Lingual-Index*, attraverso una *Top Ontology* più articolata rispetto a quella di WordNet. In primo luogo i diversi gruppi di lavoro di EuroWordNet sono giunti all'accordo di enucleare all'interno della base di dati interlinguistica un numero ristretto (poco più di un migliaio⁹¹) di nodi concettuali, detti *Common Base Concepts*, la cui struttura ontologica prescinde da fattori linguistici.

Vediamo ora i dettagli della *Top Ontology* che permette di classificare la base concettuale comune a tutti i progetti nazionali. In primo luogo è da notare che, a differenza di WordNet, la forma di classificazione privilegiata non dà luogo a

⁹⁰RODRÍGUEZ ET AL. (1998, p. 6).

⁹¹Per la precisione 1059.

6. Modelli e risorse lessico-semantiche

una struttura ad albero: infatti, anche se in WordNet è possibile associare un iponimo a più iperonimi, tuttavia i nodi della *Top Ontology* sono subordinati al massimo a un solo nodo gerarchicamente superiore. In EuroWordNet invece la *Top Ontology* è articolata in 63 tratti che sono soltanto in parte mutualmente esclusivi.

Seguendo la classificazione di Lyons⁹² tutti i nodi sono ripartiti in entità del primo, del secondo e del terzo ordine, secondo questa definizione:

- first-order entities: concrete, physical things (referred to by concrete nouns);
- second-order entities: properties, acts, processes, states, events (referred to by nouns, verbs and also adjectives);
- third-order entities: propositions which exist independently of time and space (referred to by abstract nouns).

Le entità del primo ordine sono classificate in base all'origine, alla forma, alla composizione e alla funzione. Come si può osservare:

These classes are comparable with Aristotle's Qualia roles as described in Pustejovsky's Generative lexicon, (the Agentive Role, Formal Role, Constitutional Role and Telic Role respectively: Pustejovsky, 1995) but are also based on our empirical findings to classify the BCs.⁹³

L'origine indica se l'entità deriva da un'altra entità di tipo naturale o se invece è una creazione artificiale. Nel caso di derivazione naturale da esseri viventi, è prevista una suddivisione in vegetale, animale e umano. A queste si affianca una ulteriore categoria denominata «creature» per entità soprannaturali, come le divinità antropomorfe o le creature fantastiche. Sono considerati di origine animale

Origin	Il modo in cui una entità è creata o comincia a esistere	
Natural	Ciò che non è prodotto artificialmente	ποταμός
Living	Ciò che vive e muore, oppure tessuti e fluidi organici	δάκρυμα
Plant	Vegetale	ἄμπελος
Animal	Animale	ἄετός
Human	Essere umano	ἄνθρωπος
Creature	Creatura soprannaturale	δαίμων
Artifact	Prodotto artificiale	δίφρος

⁹²LYONS (1977, p. 442–443).

⁹³RODRÍGUEZ ET AL. (1998, p. 66).

6.2 WordNet(s)

Per quanto riguarda la forma delle entità del primo ordine, si distingue fra masse di sostanze amorfe, ripartite in base allo stato della materia (solido, liquido e gassoso) e oggetti dotati di una forma determinata.

Form	La struttura di una entità fisica, geometricamente definita oppure amorfa	
Substance	Ciò che non ha una forma definita, come masse d'acqua, sabbia, etc.	
Solid	Materiale solido	ἄργυρος
Liquid	Materiale liquido	γάλα
Gas	Materiale gassose	ἄνεμος
Object	Oggetto con una forma definita	τόξον

Nel caso in cui un'entità rappresenti la parte funzionale di un'altra entità, ne sia una porzione o un frammento oppure nel caso in cui una entità sia concepita come un gruppo di membri, vengono assegnati rispettivamente i tratti «part» e «group».

Composition	Organizzazione delle entità fisiche in termini di parti, gruppi o raggruppamenti di ordine superiore	
Part	Parte di un oggetto o membro di un gruppo	ἀγή
Group	Gruppo di persone o di cose	στρατός

Il quarto tratto pertinente alle entità del primo ordine riguarda la funzione. L'elenco delle funzioni stabilite in EuroWordNet è dato nella tabella di seguito.

Function	Lo scopo, il ruolo o l'attività principale di una entità fisica.	
Vehicle	Mezzo di trasporto	ναῦς
Representation	Entità concreta usata per veicolare un messaggio	
(MoneyRepresentation)	–	–
LanguageRepresentation	Parola, testo, etc.	ἔπος
ImageRepresentation	Rappresentazione visiva	βρέτας
(Software)	–	–
Place	Entità fisica che costituisce il luogo dove si svolgono azioni o dove si trovano altre entità	γαῖα
Occupation	Persona che svolge una determinata occupazione	ἄγγελος
Instrument	Strumento (strumenti di lavoro, armi, etc.)	ἔγχος
Garment	Abiti, armature e ornamenti	πέπλος
Furniture	Arredamento	λέκτρον

6. Modelli e risorse lessico-semantiche

Covering	Copertura naturale o artificiale	καλύπτρα
Container	Contenitore o luogo ristretto e chiuso	θήκη
Comestible	Prodotto commestibile	μέλι
Building	Costruzione architettonica	δόμος

Al di là della effettiva efficacia dei tratti semantici stabiliti in EuroWordNet, ciò che sembra più utile è fare una riflessione sulla diversa struttura ottenuta rispetto a WordNet attraverso l'articolazione su quattro assi distinti delle informazioni riguardanti origine, forma, composizione e funzione.

Si prendano i versi *Pers.* 126-9:

πᾶς γὰρ ἱππηλάτας
καὶ πεδοστιβῆς λέως
σμῆνος ὡς ἐκλέλοιπεν μελισ-
σᾶν σὺν ὀρχάμῳ στρατοῦ

Le traduzioni inglesi dal *LSJ* di *σμῆνος* e di *λέως* pertinenti al nostro contesto sono rispettivamente *swarm* e *division*. In WordNet *swarm* si trova nella seguente gerarchia di iperonimi:

swarm
 animal_group
 biological_group
 group, grouping

mentre *division* ha i seguenti iperonimi:

division
 army_unit
 military_unit, military_force, military_group, force
 unit, social_unit
 organization, organisation
 social_group
 group, grouping

Al di là degli iperonimi intermedi, il paragone ha luogo fra l'*animal_group* e il *social_group*, tuttavia, per la struttura di WordNet, il rapporto fra *animal_group* e

animal da un lato e fra *social_group* e *human_being* dall'altro è stabilito attraverso la rete dei meronimi, non degli iperonimi. L'ontologia di WordNet infatti è erede dell'albero di Porfirio e in questo tipo di organizzazione le differenze specifiche non danno luogo a una tassonomia parallela a quella dei generi prossimi⁹⁴. Come si è detto sopra, WordNet permette il collegamento a più di un iperonimo, ma usa il meccanismo dell'ereditarietà multipla in modo molto ridotto. Tuttavia *animal* è collegato ad *animal_group* e *human_being* è collegato a *social_group* attraverso la relazione: *x is_member_of y*. La struttura prevalentemente ad albero dell'iponimia, combinandosi con la struttura prevalentemente ad albero della meronimia dà luogo all'organizzazione reticolare di WordNet.

EuroWordNet invece, più vicina alla semantica di scuola europea, preferisce introdurre dei tratti fra loro indipendenti che si determinano reciprocamente: in EuroWordNet uno sciame è una entità del primo livello che si riduce, alla fine della catena degli iperonimi, nei tratti *animal+group* (entità di origine animale composta in gruppo); l'ape ha il tratto *animal*, soldato il tratto *human* e una divisione dell'esercito ha i tratti *human+group*: il tratto *animal* permette quindi di raggruppare la nozione di sciame e la nozione di ape in uno stesso insieme, così come si possono raggruppare in uno stesso insieme la nozione di divisione dell'esercito e la nozione di soldato per mezzo del tratto *human*. È chiaro però che questi tratti che strutturano la *Top Ontology* di EuroWordNet, presi singolarmente, non sono essi stessi degli iperonimi: non vale infatti la relazione *x is_a_type_of y* fra *swarm* e *animal*, mentre vale fra *swarm* e *animal_group*, che si riduce ad *animal+group*.

Mentre le entità del primo ordine sono espresse da sostantivi concreti, le entità del secondo ordine possono essere espresse sia da sostantivi astratti che da verbi, da aggettivi o da avverbi. I responsabili di EuroWordNet si riferiscono alla nozione espressa dalle entità del secondo livello con il termine *Situation*, tanto per le azioni quanto per le proprietà, le relazioni, etc. Le situazioni possono essere di due tipi: dinamiche o statiche. Nel caso di situazioni dinamiche, esprimono eventi delimitati o indelimitati nel tempo, mentre nel caso di situazioni statiche espri-

⁹⁴L'argomento, ben noto, è trattato nel saggio *l'Antiporfirio* contenuto in ECO (1985).

6. Modelli e risorse lessico-semantiche

mono proprietà o relazioni. Anche in questo caso il tratto indica una pertinenza con la nozione espressa, non un rapporto di iperonimia. Ad esempio, il tratto *property* è considerato pertinente tanto a un aggettivo che indica una proprietà quanto al sostantivo astratto che nomina quella proprietà.

Situation Type	Struttura dell'evento concepito come unità concettuale provvista o priva di sviluppo nel tempo	
Dynamic	Situazione che implica una transizione da uno stato a un'altro	ἀμείβω
Bounded Event	Situazione in cui la transizione da uno stato all'altro è definita nel tempo	πλήσσω
Unbounded Event	Situazione in cui la transizione da uno stato all'altro può essere suddivisa in microtransizioni da uno stato all'altro senza soluzione di continuità	αὐξάνω
Static	Proprietà, relazione o stato di cose	κείμαι
Property	Proprietà di un ente, di una situazione, oppure valore di una proprietà	ἀγαθός
Relation	Relazione che intercorre fra due o più entità fisiche o situazioni astratte	γειτονέω

Le entità di terzo livello, le quali non sono né percepibili con i sensi (come le entità fisiche) né hanno luogo nello spazio e nel tempo (come gli eventi), sono costituite da un piccolo numero di termini astratti come *faith*, *belief*, *idea*, etc.

EuroWordNet aggiunge alle relazioni di base definite in WordNet numerose altre relazioni, ad esempio per precisare i rapporti di meronimia (parte funzionale, porzione, area, etc.), di causalità (intenzionale, non intenzionale, etc.). Per approfondimenti, si rimanda a RODRÍGUEZ ET AL. (1998).

MultiWordNet

Sviluppato in seno alla *TCC Division - Cognitive and Communication Technologies* dell'ITC-irst di Trento, MultiWordNet⁹⁵ propone un altro modello di adat-

⁹⁵I materiali e la documentazione si trovano all'indirizzo:

<http://multiwordnet.itc.it/english/home.php>.

MultiWordNet si basa sull'organizzazione di WordNet 1.6 ed è possibile interrogare la base di dati online.

tamento della struttura di WordNet in ambito multilinguistico. In EuroWordNet l'individuazione dei rapporti di iperonimia a livello microgenerico erano affidati ai diversi progetti nazionali, per giustificare il fatto che se una lingua manca di un iperonimo (*lexical gap*) è una forzatura mantenere la struttura gerarchica di un'altra lingua. In MultiWordNet invece si è preferito colmare queste lacune lessicali nelle varie lingue attraverso delle perifrasi, che del resto si trovano comunemente nei dizionari bilingui per tradurre i vocaboli che trovano una corrispondenza precisa. L'introduzione, accanto ai *synsets*, di una struttura detta *phrasets*, permette di allineare l'organizzazione della rete di WordNet alla rete concettuale di MultiWordNet, appoggiandosi all'ipotesi che la struttura concettuale è indipendente dalla lingua che la esprime: le relazioni fra i nodi concettuali rimangono le stesse e così *synsets* e *phrasets* di lingue differenti possono afferire ai medesimi nodi. Riprendendo la terminologia di Hjelmslev, i nodi della rete concettuale individuano zone della materia del contenuto alle quali si fa riferimento, in lingue diverse, o per mezzo di unità lessicali o tramite perifrasi che colmano la lacuna lessicale con la migliore approssimazione possibile.

Ad esempio la lingua inglese distingue fra i termini anatomici *finger* e *toe*, che hanno per iperonimo il vocabolo *digit*. L'italiano, privo di questa distinzione, nel tradurre, fuore contesto, *toe* con *dito* estende indebitamente il significato all'iperonimo *digit*. La frase *dito_del_piede* colma la lacuna lessicale dell'italiano e ha la medesima estensione del vocabolo da tradurre. È possibile quindi mantenere la medesima rete di relazioni fra *dito_della_mano*, *dito_del_piede* e *dito*, nonostante in italiano i concetti corrispondenti non siano lessicalizzati.

Altro punto di forza di MultiWordNet è la disambiguazione dei termini polisemici ricorrendo alla nozione di dominio (caratteristica presente anche negli altri progetti esaminati, ma non utilizzata con la stessa sistematicità). Ad esempio le cinque differenti accezioni del vocabolo *horse* sono ripartite nei seguenti domini: *zoology* per il cavallo come animale, *athletics* per la cavallina, *military* per la cavalleria, *chess* per il cavallo degli scacchi. Soltanto a *horse* nel senso di cavalletto per segare la legna non è stato attribuito ancora alcun dominio.

6.3 Ancient Greek WordNet

Con la collaborazione dei responsabili dell'*Alpheios Project*⁹⁶ si è dato l'avvio ad *Ancient Greek WordNet* (BIZZONI ET AL., 2014), tramite l'estrazione di coppie di parole Greco-Inglese dal *LSJ* e altri dizionari bilingui. Tramite il traduttore inglese si procede alla proiezione sull'originaria *Princeton WordNet*.

La risorsa ha bisogno di essere affinata in termini di precisione (*precision*), cioè di risultati rilevanti recuperati rispetto alla totalità dei valori trovati, e in termini di recupero (*recall*), cioè di risultati rilevanti recuperati rispetto a tutti i valori rilevanti, trovati o no. Ad esempio, l'attuale *Ancient Greek WordNet* non è ancora in grado di mostrare *κάτευγμα* come sinonimo di *ἀρά* nel senso di imprecazione (si veda CANDIO, 2014), perché il traduttore inglese del *LSJ* è al plurale: *imprecations*. Si sta cercando di migliorare *Ancient Greek WordNet* prendendo in considerazione l'analisi morfo-sintattica dei traduttori, come spiegato in BOSCHETTI ET AL. (2014), e combinando l'uso dei traduttori estratti dai dizionari con altre risorse lessico-semantiche, come spazi semantici costruiti su *corpora* di testi greci (BIZZONI ET AL., 2015) o il *Dynamic Lexicon* (BERTI ET AL., 2016) sviluppato presso il *Perseus Project* e l'*Open Philology Project* dell'Università di Lipsia.

A partire da *Ancient Greek WordNet* è stata creata *Homeric WordNet* (STANZIONE ET AL., 2017), focalizzata sul lessico omerico e più facile da controllare, ai fini delle correzioni manuali effettuate in collaborazione con docenti e studenti dell'Università di Pisa.

⁹⁶<http://alpheios.net>

7

Esplorazione degli spazi semantici

Questo capitolo studia gli spazi semantici basati sul *corpus* del *Thesaurus Linguae Graecae*¹. Un *corpus* diacronico costituito da testi letterari antichi è di dimensioni relativamente piccole (non più di alcune centinaia di *megabytes*), se paragonato a *corpora* moderni, letterari e non letterari (fino ad alcuni *terabytes*). È formato da testi che si distribuiscono su più secoli, rendendo interessante lo studio del cambiamento semantico dovuto a mutamenti culturali, scientifici e spirituali. Infine, non potendo essere valutato da parlanti nativi in grado di cogliere le originarie sfumature di significato, rende particolarmente utile lo studio dei contesti d'uso delle singole parole esaminate.

¹Un approccio esemplare basato sul *corpus* è stato applicato da O'DONNELL (2005), che focalizza la sua attenzione sul Greco del Nuovo Testamento. In particolare, l'autore affronta il tema delle varianti testuali e dei criteri basati su *corpora* testuali per selezionarle. Gran parte del suo lavoro è dedicato alle analisi lessicografiche del Nuovo Testamento. Lo studio delle collocazioni mostra, ad esempio, come *ἐγείρω* (svegliare) e *ἀνίστημι* (salire) siano sinonimi nello specifico contesto della risurrezione cristiana. Questo capitolo amplia e aggiorna quanto espresso in BOSCHETTI (2009).

7.1 Struttura, costruzione e proprietà degli spazi semantici

Uno spazio semantico è un modello computazionale basato sulla distribuzione delle parole di un *corpus*, allo scopo di rappresentare la similarità semantica tramite la prossimità in spazi vettoriali (FABRE E LENCI, 2015; LENCI, 2008; SAHLGREN, 2006, 2008; SALTON ET AL., 1975). Lo spazio multidimensionale dove sono collocati i termini è costruito tramite una tabella di co-occorrenza (Tabella 7.1), che incrocia le m parole lemmatizzate del *corpus* con le n parole lemmatizzate che appaiono nella finestra di contesto (di larghezza fissa w) della parola sotto osservazione. Il valore di ciascuna cella esprime il numero di volte che il termine nella riga co-occorre all'interno della finestra di contesto con il termine nella colonna. Ad esempio, la cella che incrocia *ἀβρόγος* con *ἀβροχίτων* ha valore 1 perché le due parole co-occorrono soltanto una volta nella finestra di 20 parole (rispettivamente ai versi *Pers.* 541 e 543). Al contrario, *ἀβροδαίτος* non appare mai vicino a questi termini nella medesima finestra di contesto. La larghezza della finestra di contesto, w , può essere allargata o ristretta. Scelte comuni sono dieci, venti o cento parole a sinistra e a destra della parola presa in esame (si veda SAHLGREN, 2006, p. 68). Una finestra stretta mette in luce le relazioni sintagmatiche, mentre una finestra più larga mette in luce le relazioni paradigmatiche. L'iperspazio costruito con i dati della tabella di co-occorrenza ha n dimensioni e contiene m punti. Le celle contengono le coordinate di ciascun punto nello spazio multidimensionale. In questo modo, punti vicini nell'iperspazio rappresentano termini che appartengono a contesti simili. Infatti, la teoria degli spazi semantici è basata sull'ipotesi distribuzionale: “Words that are similar in meaning occur in similar contexts” (RUBENSTEIN E GOODENOUGH, 1965).

L'applicazione della Singular Value Decomposition (WALL ET AL., 2003) permette la riduzione delle dimensioni dello spazio semantico originale, riducendo il rumore (*noise*) nei dati. Grazie alla riduzione delle dimensioni, termini che non occorrono mai nella stessa finestra di contesto ma che occorrono con termini che a loro volta sono semanticamente simili, diventano vicini nel nuovo spazio a di-

7.1 Struttura, costruzione e proprietà degli spazi semantici

	...	<i>ἀβρόγος</i>	<i>ἀβροδίατος</i>	...	<i>ἀβροχίτων</i>	...
...
<i>ἀβρόγος</i>	...	0	0	...	1	...
<i>ἀβροδίατος</i>	...	0	0	...	0	...
...
<i>ἀβροχίτων</i>	...	1	0	...	0	...
...

Tabella 7.1: Tavola delle co-occorrenze

mensioni ridotte. Secondo GROSSMAN E FRIEDER (2004), “The key to similarity is not that two terms happen to occur in the same document [or window]: it is that two terms appear in the same *context*, - that is they have very similar neighboring terms”. Qui il termine *contesto* è usato nel senso ampio di co-occorrenza del primo ordine o del secondo ordine. Infatti, se AB co-occorre nel contesto x , BC nel contesto y e CD nel contesto z (co-occorrenze del primo ordine), AC e BD sono co-occorrenze del secondo ordine e AD è una co-occorrenza del terzo ordine (KONTOSTATHIS E POTTENGER, 2003). La Singular Value Decomposition permette di catturare queste relazioni. Le nuove dimensioni dell’iperspazio ridotto non corrispondono più alle singole parole della finestra di contesto: le co-occorrenze di ordine superiore emergono dal raggruppamento di informazioni distribuite nelle dimensioni originali.

Rispetto ad altri modelli computazionali usati per determinare la similarità semantica tra termini (si veda ad esempio FELLBAUM, 1998a), in questo modello “the space is constructed with no human intervention, and with no a priori knowledge or constraints about meaning similarities” (SAHLGREN, 2006, p. 21).

I metodi applicati in questo studio sono usati dai linguisti in ampi *corpora* di documenti non letterari, in Inglese o in altre lingue moderne, su molti milioni di parole. Come si spiegherà più oltre, il numero di occorrenze influenza pesantemente l’individuazione di relazioni semantiche rilevanti.

7. Esplorazione degli spazi semantici

Corpora molto ampi in lingua inglese possono essere analizzati anche senza lemmatizzazione. Al contrario, data la complessità morfologica del Greco e la relativamente piccola dimensione del *corpus* del TLG (approssimativamente 76 milioni di parole, da cui sono state escluse circa quattro milioni di parole appartenenti a opere frammentarie), la lemmatizzazione si rende necessaria, perché riduce la rarefazione dei dati. Le forme flesse sono state lemmatizzate con *Morpheus* (CRANE, 1991), che è il lemmatizzatore più accurato per il Greco antico.

Uno dei problemi principali è dovuto alle parole composte, perché sono lemmatizzate senza alcuna analisi delle parti che la compongono, perdendo quindi la relazione semantica. Ad esempio, le occorrenze di *ἀβρο-* (con l'idea di lusso, dolcezza, delicatezza, fascino) non sono registrate in *Pers.* 1073 *ἀβροβάτης* (che incede delicatamente), *Pers.* 541 *ἀβρόγυος* (pianto delicato), *Pers.* 135 *ἀβροπενθής* (secondo alcune interpretazioni: delicatamente sofferente, secondo altre: sofferente in modo vistoso, quasi sofisticato) e *Pers.* 543 *ἀβροχίτων* (delicatamente rivestito).

In questo studio sono stati generati quattro spazi semantici a partire dal TLG: lo spazio dei poemi omerici, il *corpus* della tragedia greca con i relativi *scholia* e il *corpus* filosofico. La selezione dei testi si è basata sui generi letterari, come indicato in SQUITIER (1990).

Le parole sono state elaborate con *Infomap*², un'applicazione per lo studio degli spazi semantici³, che genera tabelle di co-occorrenze, esegue la Singular Value Decomposition e calcola le coordinate dei termini nello spazio semantico risultante. I parametri sono stati impostati a 120.000 righe, 2.000 colonne, 300 valori singolari, 300 iterazioni per la scomposizione dei valori singolari, 100 parole di parole di contesto a sinistra e 100 parole di contesto a destra.

Per ogni termine è possibile ottenere la lista dei termini più vicini nello spazio semantico e quindi, secondo l'ipotesi distribuzionale, i termini semanticamente più simili alla parola sotto osservazione. Ad esempio, dato il termine *κυβερνήτης* (nocchiero) e lo spazio semantico dell'intero TLG, le prime dieci associazioni se-

²<http://infomap-nlp.sourceforge.net>

³Si veda PENNINGTON ET AL. (2014) per lo stato dell'arte degli strumenti correntemente usati.

7.1 Struttura, costruzione e proprietà degli spazi semantici

mantiche ordinate per similarità sono: *ναύτης* (marinaio) 0.68, *κυβερνάω* (governo la nave) 0.61, *σκάφος* (scafo) 0.57, *πλοῖον* (nave) 0.55, *πρύμνα* (poppa) 0.53, *πηδάλιον* (timone) 0.52, *ἄγκυρα* (ancora) 0.51, *κῦμα* (onda) 0.50, *κυβερνητικός* (adatto a governare la nave) 0.50, *πλωτήρ* (navigabile) 0.50. La *cosine similarity*, calcolata da *Infomap*, è il prodotto scalare dei vettori della frequenza normalizzata. Data una o più parole, uno strumento di similarità offerto da *Infomap*, chiamato *associate*, offre una lista di termini fra loro associati e ordinati in base alla *cosine similarity*.

Le relazioni semantiche sono distribuite sia sull'asse sintagmatico che sull'asse paradigmatico: sinonimia, come *θάλασσα* (mare) - *πέλαγος* (mare aperto); iperonimia, iponimia, co-iponimia, come *ὄρνις* (uccello) - *ίέραξ* (falco) - *ἀετός* (aquila); olonimia, meronimia, co-meronimia, come *ναῦς* (nave) - *πρύμνα* (poppa) - *ἰστίον* (vela); antonimia, come *μέγας* (grande) - *μικρός* (piccolo) o *θάλασσα* (mare) - *ἡπειρος* (terra ferma); appartenenza allo stesso frame, secondo la definizione di FILLMORE (1997), come *ἵππος* (cavallo) - *ἄρμα* (carro) o *θάλασσα* (mare) - *ναύτης* (marinaio).

La frequenza delle parole ha una forte influenza sulle associazioni semantiche. In effetti, la parola sotto osservazione deve apparire in un numero adatto di contesti. I risultati sperimentali emersi in questo studio suggeriscono che è necessario avere almeno quattro occorrenze del termine, altrimenti il sistema trova associazioni incoerenti. Tuttavia, a causa della legge di Zipf (EVERT E BARONI, 2007; ZIPF, 1949), in un relativamente piccolo *corpus* testuale come il TLG, più della metà delle parole compaiono meno di quattro volte. Di conseguenza, i risultati sono soddisfacenti soltanto per una parte limitata del *corpus*. Ad esempio, considerando una parola ad alta frequenza come *λίθος* (pietra), con 14.830 occorrenze, tutte le prime dieci associazioni sono coerenti: gli iponimi *ἰάκινθος* (acquamarina) e *σμάραγδος* (smeraldo); i coiponimi *σίδηρον* (ferro) e *ξύλον* (legno), con l'aggettivo *χάλκεος* (di bronzo), derivato da un coiponimo; i meronimi *τοιχος* (muro), *ὄροφος* (tetto)⁴, *κίων* (colonna) e *ἄγαλμα* (statua).

⁴*ὄροφος* significa “canna usata per case con tetto di paglia” e, metonimicamente, tetto (*ὄροφή*).

7. Esplorazione degli spazi semantici

Come detto sopra, le associazioni semantiche relative a termini a bassa frequenza sono inaffidabili. Infatti, questi termini (specialmente termini con una singola occorrenza) offrono associazioni con altri termini a bassa frequenza nello stesso contesto, senza coerenza semantica. Ad esempio, *ἀργυροφειγής*, brillante come l'argento, occorre solo quattro volte nel *corpus* del TLG, e perfino le prime quattro associazioni sono prive di relazione semantica evidente: *ἀρχοντιάω* (desidero il comando), *μεθυσφάλέω* (sono ubriaco) *ὁμόφθογγος* (che suona insieme), *περισσόνοος* (eminente nella comprensione).

Ma in particolari condizioni anche le associazioni relative a parole a bassa frequenza possono essere significative. Ad esempio quando il termine occorre una volta nel contesto originale e alcune altre volte in un'opera di lessicografia o in uno *scholium*. Ad esempio *ἀβρόγος*, lamento femminile, occorre una sola volta in Tragedia (*Pers.* 541 *ἀβρόγος*), ma tre volte negli *scholia*. Nello spazio semantico degli *scholia*, il termine produce le seguenti associazioni rilevanti: *ὑγιοζυγία* (unione sana), *ἀρτιζυγία*, (unione recente, che nel contesto di questa tragedia è la causa del pianto), *κατερείκω* (strappo), *ἀβροχίτων* (con vesti morbide), *πρωτόμορος* (che muore prima), *κατασχίζω* (piango), *ἀπαλόχροος* (con la pelle morbida). Quattro di questi termini sono parole a bassa frequenza che appaiono nella stessa finestra di contesto: *Pers.* 542 *ἀρτιζυγίαν* (*ἀρτιζυγία*), 538 *κατερεικόμεναι* (*κατερείκω*), 568 *πρωτομόροιο* (*πρωτόμορος*) e 543 *ἀβροχίτωνας* (*ἀβροχίτων*). Due termini sono *glossae*: *ὑγιοζυγία* spiega *ἀρτιζυγία* e *κατασχίζω* spiega *κατερείκω*. Infine, il primo componente del composto *ἀπαλόχροος*, *ἀπαλός* (delicato), è un sinonimo di *ἀβρός*, il primo componente di *ἀβρόγος*.

7.2 Osservazioni sui cambiamenti diacronici di significato

La natura diacronica del *corpus* di Greco antico⁵ mostra la sua influenza nelle associazioni, specialmente per quanto riguarda i termini astratti. Il *corpus* del

⁵Lo studio del *corpus* di testi greci tramite semantica distribuzionale in diacronia è stato ripreso in modo approfondito da RODDA ET AL. (2017).

7.2 Osservazioni sui cambiamenti diacronici di significato

TLG è stato diviso in due parti: la prima contiene i testi prodotti prima dell'era volgare e la seconda quelli prodotti dopo. Sono stati scelti arbitrariamente due termini: uno concreto, *θάλασσα* (mare) e uno astratto, *θάνατος* (morte).

<i>θάλασσα</i>	mare	<i>θάλασσα</i>	mare
<i>πέλαγος</i>	alto mare	<i>πέλαγος</i>	alto mare
<i>ἤπειρος</i>	terra	<i>ἤπειρος</i>	terra
<i>ποταμός</i>	fiume	<i>λίμνη</i>	palude
<i>νῆσος</i>	isola	<i>κῶμα</i>	onda
<i>πότιμος</i>	potabile	<i>ποταμός</i>	fiume
<i>λίμνη</i>	palude	<i>πότιμος</i>	potabile
<i>άλμυρός</i>	sale	<i>ἐπικλύζω</i>	trabocco
<i>πόντος</i>	mare aperto	<i>άλμυρός</i>	sale
<i>ὄρος</i>	monte	<i>ἀπειρώω</i>	vado verso l'infinito
<i>κῶμα</i>	wave	<i>Ὠκεανός</i>	Ocean

Tabella 7.2: *θάλασσα*: associazioni prima e dopo l'era volgare

Come mostra la Tabella 7.2, *θάλασσα* genera associazioni con termini geografici o aggettivi e verbi correlati, *θάνατος* offre due differenti liste che dimostrano il cambiamento culturale e le connotazioni religiose. Perciò, la seconda lista contiene termini con una forte valenza cristiana: *ἄδης* (gli inferi, ma nell'era volgare: l'inferno cristiano), *νεκρώω* (mortifico), *ἀνάστασις* (resurrezione), *ἀθανασία* (immortalità), *Θάνατος* (la personificazione della morte) e, infine, il termine cristiano più esplicito: *συσταυρόομαι* (sono crocifisso insieme).

Nel caso di termini con significato letterale e figurato, il significato usato più frequentemente nel *corpus* (o in una partizione del *corpus*) oscura il significato secondario. Ad esempio, *ναύτης* (marinaio) occorre 1,161 volte nel TLG, principalmente nel suo senso letterale. Di conseguenza, le principali associazioni semantiche sono: *κυβερνήτης* (nocchiero), *πλοῖον* (nave), *σκάφος* (scafo), *ναῦς* (nave), *πρύμνα* (poppa), *κῶμα* (onda), *ἄγκυρα* (ancora), *πέλαγος* (mare aperto), *ἔμπορος* (passeggero di una nave), *ἐμπλέω* (navigo). Tuttavia, nello spazio semantico costruito sul *corpus* delle opere filosofiche contenute nel TLG, *ναύτης* è

7. Esplorazione degli spazi semantici

θάνατος	morte	θάνατος	morte
ἀποθνήσκω	muoio	ἀποθνήσκω	morire
ἀποκτείνω	uccido	νεκρός	cadavere
τιμωρία	vendetta	ᾗδης	inferno
φόνος	strage	νεκρώω	mortifico
φεύγω	fuggo	ἀνάστασις	resurrezione
θνήσκω	muoio	ἀθανασία	immortalità
συμφορά	disgrazia	θνήσκω	muoio
ἀνδροφόνος	omicidio	Θάνατος	Morte
τραῦμα	ferita	συσταυρόμαι	sono crocifisso insieme
τελευτάω	porto a termine	ζωή	vita

Tabella 7.3: θάνατος: associazioni prima e dopo Cristo

usato solo 79 volte (17 volte solo nelle opere di Platone), principalmente come una metafora nel dominio dell'attività politica. In questo caso tra le associazioni più pertinenti troviamo: ἄρχω (governare), ὑπήκοος (alleato obbediente), στρατηγέω (sono un generale), μονοπώλιον (diritto di monopolio), παραβοηθέω (vengo ad aiutare gli alleati).

Mentre l'intero *corpus* oscura i significati secondari dei termini polisemici, le partizioni del *corpus* secondo criteri tematici (ad esempio: testi filosofici, grammaticali, medici, etc.) sono in grado di far emergere i domini specifici di significato. Ad esempio, il termine polisemico πούς (piede), nello spazio semantico dell'intero TLG produce la lista ποδώ (sono provvisto di piedi), ποδῖς (scarpa), δάκτυλος (dito), σκοῦτα (scudo), σκέλος (gamba), παρασκελής (con gambe disuguali), ὄργυια (la lunghezza delle braccia protese) e βησαλικόν (opera in mattoni) omettendo diversi domini di significato. Al contrario, nello spazio semantico della medicina lo stesso termine produce il seguente elenco di associazioni: ποδῖς (scarpa), πτέρνα (calcagno), σκέλος (gamba), χεῖρ (mano), ποδώ (sono provvisto di piedi), κνήμη (parte tra ginocchio e caviglia), γόνυ (ginocchio), πεδάω (lego), δάκτυλος (dito) e κάμπτω (piegare). Come si vede, la maggior parte di questi termini sono relativi al dominio delle parti del corpo. Nello spazio semantico

degli *scholia* lo stesso termine produce la lista: ἴαμβος (giambo), τροχαικός (trocaico), δάκτυλος (dattilo), ὑπερκατάληκτος (ipercatalettico), τροχαιῖος, ἀδιάφορος (indifferente da un punto di vista metrico), βραχυκατάληκτος (che termina con una sillaba breve), ἀντίσπαστος (antispasto) e παιών (peana), tutti relativi al dominio della metrica.

Lo strumento *associate* può ricevere come parametro un singolo termine o una lista di termini. Nel caso di termini polisemici, fornire una lista di parametri può determinare uno specifico dominio d'indagine, anche se lo spazio semantico è costruito sull'intero TLG. Ad esempio, usare come chiave di ricerca ποῦς (piede) e χεῖρ (mano) permette di trovare una serie di associazioni relative alle parti del corpo, mentre usare ποῦς (piede) e ἴαμβος (giambo) permette di trovare associazioni nel campo dell'analisi metrica.

7.3 Relazioni semantiche osservate

La riduzione delle dimensioni dello spazio semantico originario permette la rappresentazione delle distanze fra termini in un grafo bidimensionale. La riduzione delle dimensioni viene eseguita in due fasi: nella prima fase *Infomap* riduce le dimensioni originali, che sono più di un migliaio, a 300, grazie alla scomposizione in valori singolari (SVD). Nella seconda fase, viene eseguito il *multidimensional scaling* tramite uno script in R.

Per rappresentare i dati in un grafico bidimensionale, dopo avere effettuato il *multidimensional scaling* (BAAYEN, 2008, pp. 146–148), si procede alla partizione dei gruppi di termini trovati tramite il metodo k-means (si veda FENG E MANMATHAN, 2006).

Questo metodo di raggruppamento è particolarmente efficace per gli antonomi e per i termini strutturabili gerarchicamente. I membri di una coppia antonomica, condividendo contesti molto simili, tendono ad essere raggruppati strettamente fra di loro e a essere invece separati dalle altre coppie. I membri di tassonomie tendono a essere disposti in base alla gerarchia tassonomica e a formare gruppi per ciascuna categoria con sottogruppi organizzati in modo significativo.

7. Esplorazione degli spazi semantici

Gli antonimi stabiliscono sia relazioni sintagmatiche che paradigmatiche. Sul piano sintagmatico, coppie di antonimi appaiono frequentemente nello stesso contesto. Specialmente in testi filosofici e in opere metalinguistiche, come grammatiche e lessici, commentari e *scholia*, coppie di antonimi sono oggetto di studio.

Sull'asse paradigmatico, gli antonimi compaiono in contesti simili, perché esprimono proprietà opposte degli stessi termini, in particolare quando la coppia di antonimi è costituita da aggettivi che possono occorrere con un piccolo numero di sostantivi.

Per questi motivi, ci si aspetta che i componenti delle coppie antonimiche siano molto vicini nello spazio semantico ridotto. Si osservano relazioni strette per termini di un dominio specifico, come *ἔμψυχος* (animato) e *ἄψυχος* (inanimato), comunemente usati in testi filosofici e medici, anche se il senso metaforico può essere correlato a un numero limitato di altri domini, come la retorica (ad esempio, un discorso animato o uno stile senz'anima). Rapporti meno stretti sono osservati per coppie di contrari che possono essere usati in qualsiasi contesto, come *μέγας* e *μικρός*. Ad esempio, il valore di similarità di *ἄψυχος* rispetto a *ἔμψυχος* è 0.8, mentre il valore di similarità di *μικρός* rispetto a *μέγας* è 0.6. Il valore di similarità è fornito sempre dallo strumento *associate*.

Infine, vale la pena notare che le parole polisemiche possono avere contrari differenti, a seconda dei diversi sensi (si veda FELLBAUM, 1998a, p. 51). Ad esempio, *βαρύς* nel senso di “grave”, ha come antonimo *δξύς* nel senso di “acuto”, ma *βαρύς*, nel senso di “pesante” ha *κοῦφος* fra i suoi antonimi, nel senso di “leggero”.

L'esempio presentato in Figura 7.1 si basa su nove coppie di antonimi selezionati arbitrariamente ed elencati in Tabella 7.4. Tutte le parole sono termini ad alta frequenza: il più frequente è *ἄνθρωπος* (uomo) con 51845 occorrenze e il meno frequente è *ἄψυχος* (inanimato) con 2712 occorrenze. Come mostrato nella figura, tutte le coppie sono raggruppate correttamente usando il metodo di raggruppamento k-means.

7.3 Relazioni semantiche osservate

ἀνήρ	uomo	γυνή	donna
ἄρσην	maschile	θηλυς	femminile
γλυκός	dolce	πικρός	amaro
ἔμψυχος	animato	ἄψυχος	inanimato
ἡμερος	giorno	νύξ	notte
θερμός	caldo	ψυχρός	freddo
λογικός	razionale	ἄλογος	irrazionale
ὀξύς	acuto	βαρύς	grave
φάος	luce	σκότος	tenebra

Tabella 7.4: Antonimia

La coppie sembrano disposte secondo una struttura gerarchica. Il dendrogramma in Figura 7.2, che è solo una rappresentazione grafica diversa degli stessi dati (sempre basata su due dimensioni), evidenzia meglio questa struttura.

Una semplice interpretazione di questi dati suggerisce che ἄρσην (maschio) e θηλυς (femmina) siano raggruppati con ἀνήρ (uomo) e γυνή (donna) come determinazioni di genere dell'iperonimo implicito ἄνθρωπος (essere umano). λογικός (razionale) e ἄλογος (irrazionale) sono implicati da ἔμψυχος, il termine positivo della coppia antonimica ἔμψυχος (animato) e ἄψυχος (inanimato) mentre λογικός (razionale) rappresenta la differenza specifica, in termini aristotelici, di ἄνθρωπος (essere umano). Nella seconda metà del grafico, φάος (luce) e σκότος (oscurità) sono raggruppati con ἡμερος (giorno) e νύξ (notte) a causa della forte associazione tra questi termini. γλυκός (dolce) e πικρός (amaro), θερμός (caldo) e ψυχρός (freddo), e infine ὀξύς (acuto) e βαρύς (grave), sono raggruppati come percezioni gustative, tattili e auditive, collegate al precedente gruppo di parole relative alle percezioni visive.

7. Esplorazione degli spazi semantici

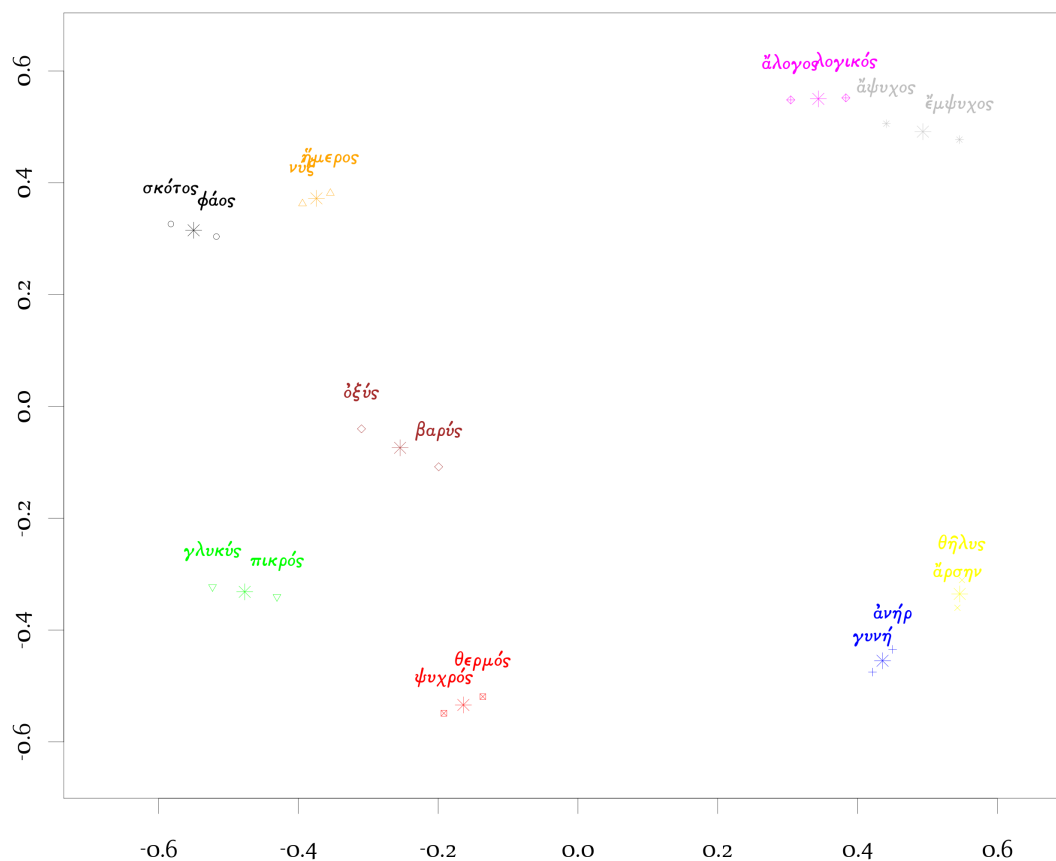


Figura 7.1: Raggruppamenti di antonimi

Tassonomie

Anche se la dimensione del TLG non è grande, le tassonomie di maggior rilievo possono essere facilmente identificate e gli elementi classificati in modo apparentemente errato, a volte possono essere spiegati grazie a sensi secondari dei termini.

Quando i termini presi in esame appartengano a paradigmi ben consolidati, i risultati sono soddisfacenti sia per significati concreti, come parti del corpo, parti della nave, parti di edifici, sia per significati astratti, come relazioni familiari, emozioni o virtù.

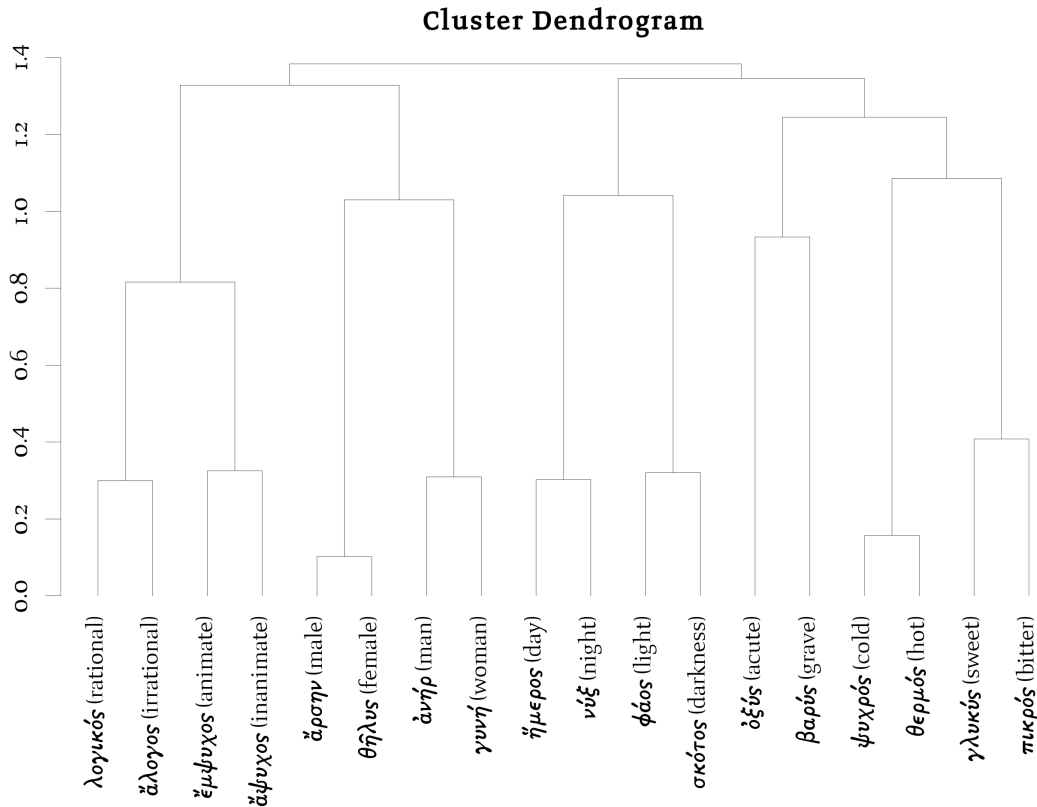


Figura 7.2: Gerarchia di antonimi

In Figura 7.3, basata sui dati di Tabella 7.5, viene mostrato come i termini della tassonomia animale siano disposti nel grafico bidimensionale. La partizione tradizionale in terra, con *χερσαῖος* (terrestre) o *πεζός* (a piedi), acqua, con *ἔνυδρος* (acquatico) e aria, con *πτηνός* (volatile) è conservata, anche se alcuni casi devono essere discussi. Il significato principale di *λαγῶς* è la lepre, un animale terrestre, ma il termine è classificato come uccello. La classificazione errata è solo apparente perché *λαγῶς*, come indicato nel LSJ, è anche “a bird with rough feathered feet, mentioned with the swallow”. Inoltre, è anche “a kind of sea-slug, *Lepus marinus*”. Allo stesso modo, *στρουθός* (passero) è un uccello ma anche “a flat fish, flounder, *Pleuronectes flesus*”. Come mostrato nel grafico, *στρουθός* si trova nel mezzo tra uccelli e pesci. Gli iperonimi *θήρ* (bestia), *ὄρνις* (uccello) e *ἰχθύς*

7. Esplorazione degli spazi semantici

θήρ	bestia	ὄρνις	uccello	ἰχθύς	pesce
κρίος	montone	γύψ	avvoltoio	τρίγλη	triglia
κύων	cane	χελιδών	rondine	σκάρος	pesce pappagallo
βοῦς	bue	ἀετός	aquila	θύννος	tonno
ῥῖς	maiale	πέρδιξ	pernice	λάβραξ	spigola
λέων	leone	ἀλεκτρυών	gallo	κέφαλος	cefalo
ἔλαφος	cervo	κύκνος	cigno	ἔγχελυς	anguilla
ταῦρος	toro	τέττιξ	cicala		
ὄνος	asino	ἰέραξ	falco		
ἵππος	cavallo	στρουθός	passero		
ἡμίονος	mulo				
λαγῶς	lepre				

Tabella 7.5: Animali

(pesce) sono al centro del grafico. L'insetto volante *τέττιξ* (cicala) è classificato come un animale che vola, quindi insieme agli uccelli.

Nell'esempio precedente, i termini sono stati selezionati arbitrariamente. L'aggiunta, la sottrazione o la sostituzione di parole può modificare i rapporti reciproci fra tali parole, producendo delle classificazioni errate. Nell'esempio corrente, la modifica dei termini può produrre, in particolare, sovrapposizioni tra bestie e uccelli, per la frequente condivisione di caratteristiche (ad esempio rapidità o intelligenza), mentre la classe dei pesci tende a evitare sovrapposizioni con le altre classi.

Per ridurre l'arbitrarietà nella creazione dell'elenco, è possibile utilizzare delle cosiddette parole-seme (*seed words*), cioè delle parole fortemente prototipiche, per creare la lista di parole per ciascun raggruppamento previsto, prendendo le prime *n* parole più simili alle parole-seme.

Lo strumento *associate* individua similarità semantiche relative a singole parole ma anche a gruppi di parole. Una coppia di parole è sufficiente per generare una lista di associazioni coerenti. Se le coppie di parole usate come semi sono suffi-

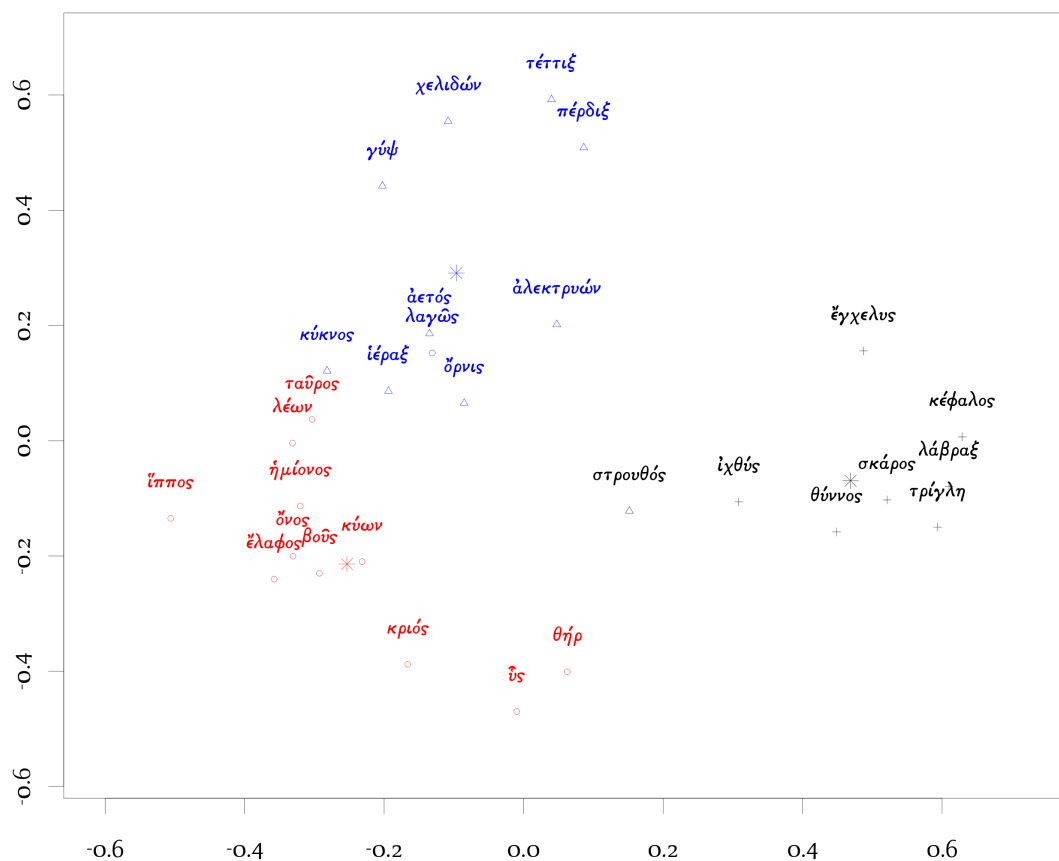


Figura 7.3: Clusters di animali

cientemente contrastive, ci si aspetta che le liste di parole generate si raggruppino senza sovrapposizioni. La regola empirica per scegliere parole-seme significative è quella di osservare i grafici basati su liste di parole arbitrarie, come visto sopra, e poi selezionare una coppia di termini per ciascun raggruppamento che stia ai margini, facendo attenzione che almeno uno dei termini sia lontano dall'area di sovrapposizione fra i raggruppamenti.

Un esempio dell'applicazione di questo metodo è mostrato in Figura 7.4. In questo caso, le coppie di parole usate come seme sono state selezionate dal grafico precedente, generato con elementi arbitrari (Figura 7.3). *ἵππος* (cavallo) e *κύνων* (cane) sono stati selezionati per gli animali terrestri, *ἰέραξ* (falco) e *πέρδιξ* (pernice) per i volatili e, infine, *ἔγχελυς* (anguilla) e *θύννος* (tonno) per gli ani-

7. Esplorazione degli spazi semantici

mali acquatici. I risultati forniti dallo strumento *associate* per ogni coppia di parole-seme sono mostrati in Tabella 7.6.

κύων	cane	ἰέραξ	falco	ἔγχυλος	anguilla
λέων	leone	πέρδιξ	pernice	θύννος	tonno
λύκος	lupo	ὄρνις	uccello	κωβίος	ghiozzo
ἔλαφος	cervo	χελιδών	passero	γαλέος	pescecane
πάρδαλις	leopardo	άλώπηξ	volpe	ἰχθύδιον	pesciolino
άλώπηξ	volpe	πτηνός	volatile	εὔχυλος	succulento
θήρα	caccia	ἀλεκτρυών	gallo	κέφαλος	cefalo
ἵππος	cavallo	ἀετός	aquila	θυννίς	tonno femmina
σκύμνος	cucciolo di leone	λαγῶς	lepre	τίφη	imbarcazione
κυνηγός	segugio	ἰχθύδιον	pesciolino	κίχλη	pesce di mare

Tabella 7.6: Lista di animali generata tramite parole-seme

Nella prima colonna, che riguarda animali terrestri, si trovano una serie di co-ponimi delle parole-seme ἵππος (cavallo) e κύων (cane). θήρα (caccia alle bestie selvagge) appartiene al *frame* della caccia, quindi agli animali terrestri. La parola condivide anche alcune forme flesse con l'iperonimo θήρ (bestia): θήρ' e θηρών. *Morpheus*, l'analizzatore morfologico, assegnando entrambe le forme a θήρ, fa perdere per quelle occorrenze l'associazione con θήρα.

Nella seconda colonna, che riguarda volatili, si trovano non solo co-ponimi di ἰέραξ (falco) e πέρδιξ (pernice), ma anche l'iperonimo ὄρνις (uccello) e l'iperonimo di questo, πτηνός (animale alato), che include uccelli e insetti volanti. λαγῶς, come visto sopra, significa "lepre" ma anche un tipo di uccello, mentre άλώπηξ (volpe), presente nella lista delle bestie terrestri, indica anche un grande pipistrello. ἰχθύδιον (piccolo pesce), nutrimento degli uccelli marini, è presente sia nella lista degli uccelli che nella lista dei pesci, ma nel grafico è classificato correttamente come pesce.

7.3 Relazioni semantiche osservate

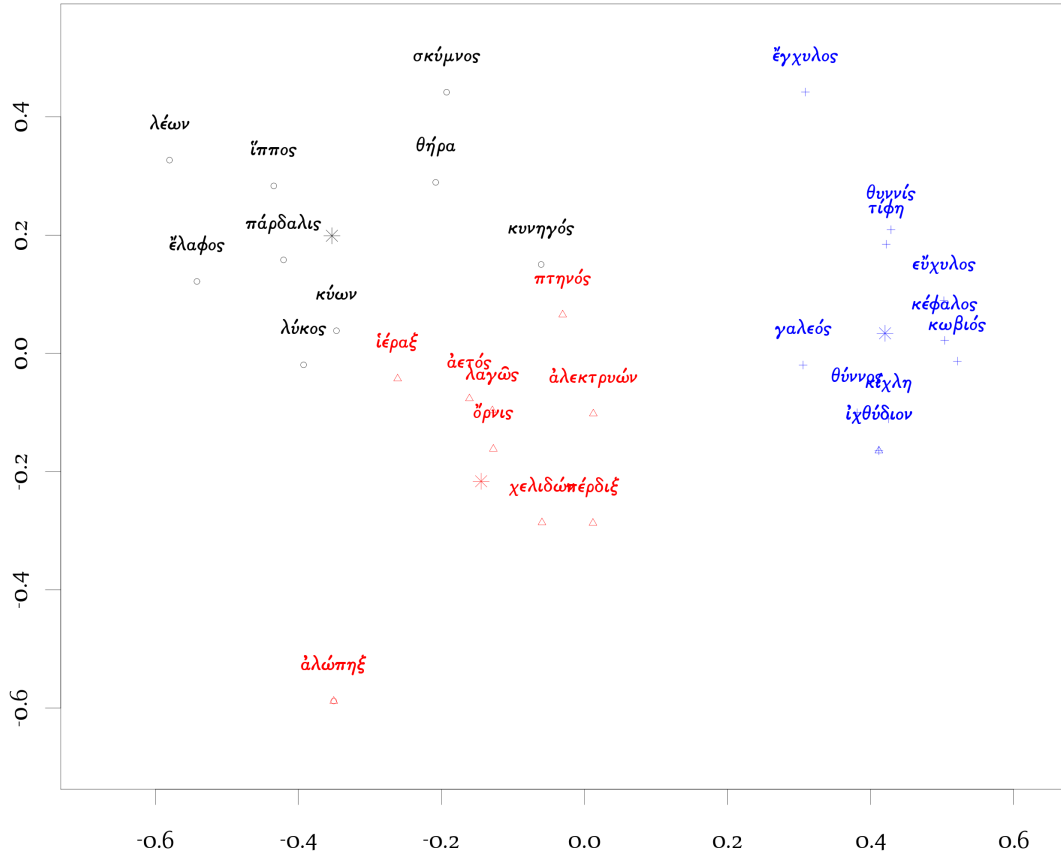


Figura 7.4: Clusters di animali generati tramite parole-seme

Infine, nella terza colonna si trovano coiponimi di ἔγχελυς (anguilla) e θύννος (tonno), ma anche εὐχυλος (succulento), un attributo applicabile al pesce cotto e τίφη, un tipo di insetto ma anche un tipo di imbarcazione.

Il dendrogramma generato su questi dati non ha la stessa coerenza osservata nel dendrogramma generato tramite termini scelti arbitrariamente, ma alcuni raggruppamenti mostrano dei *patterns* interessanti. In particolare, la coppia ἀετός (aquila) – λαγῶς (lepre) e πάρδαλις (leopardo) – ἐλάφος (cervo) mostra una forte associazione fra predatore e preda. La tripletta σκύμνος (cucciolo di leone) – θήρα (caccia) – κυνηγός (segugio, cacciatore) conferma l'idea che, all'interno delle categorizzazioni del grafico in Figura 7.5, le relazioni sintagmatiche espresse in

7. Esplorazione degli spazi semantici

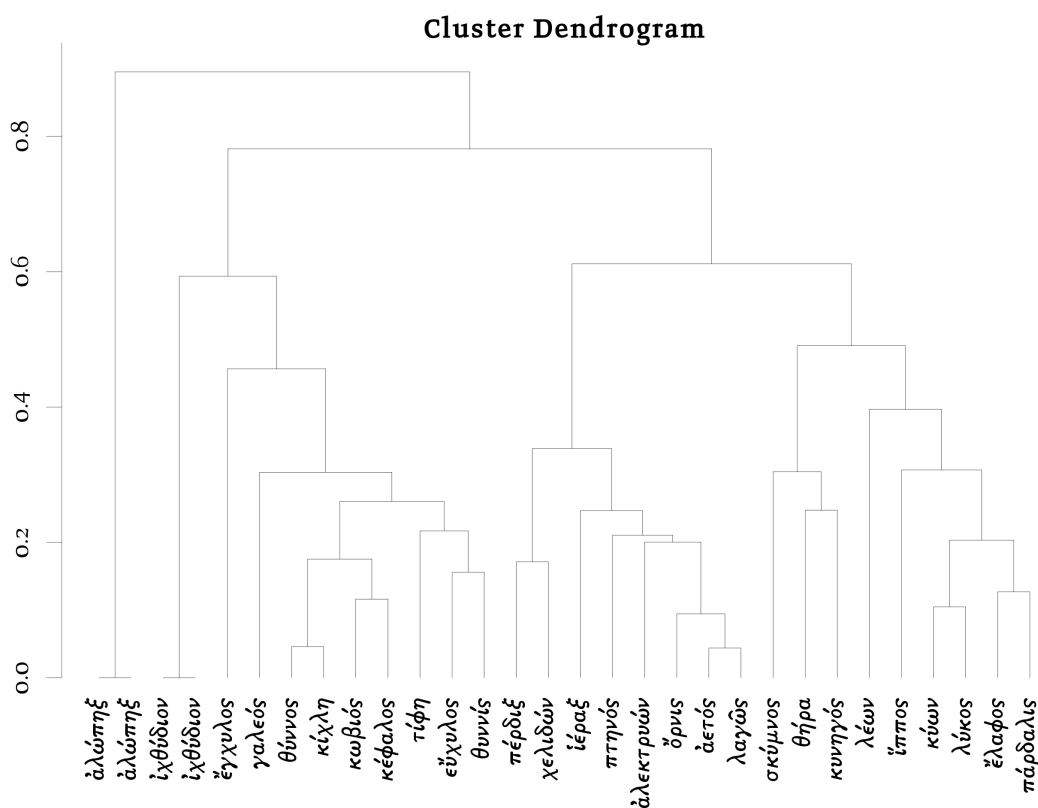


Figura 7.5: Dendrogramma di animali generato tramite parole-seme

frames e scene tipiche (ad esempio: la caccia) possa essere catturata tramite una diversa visualizzazione degli stessi dati.

Le parole-seme sono particolarmente efficaci per generare elenchi di co-meronimi. Tra i domini più produttivi, corrispondenti alle conoscenze tecniche e scientifiche degli antichi greci, sono stati testati con successo: il dominio botanico (usando i semi *φύλλον* (foglia) e *κλάδος* (ramo): parti della pianta); il dominio della medicina (usando *λάρυγξ* (laringe) e *στόμα* (bocca): parti del corpo); il dominio della tecnica nautica (usando *πρύμνα* (poppa) e *πρῶρα* (prua): parti della nave); tecnica militare (usando *κνημίς* (schiniere) e *ἀσπίς* (scudo): parti dell'armatura) e tecnica di costruzione (usando *κίων* (pilastro) e *αὐλή* (cortile): parti dell'edificio).

7.3 Relazioni semantiche osservate

La maggior parte delle cinquanta parole generate da questi semi sono meronimi e sono raggruppati senza sovrapposizioni rilevanti.

Come anticipato nell'introduzione di questa sezione, i termini astratti che appartengono a tavole tradizionali, come ad esempio tavole di virtù, sentimenti o emozioni, possono essere classificati correttamente. La tabella 7.7 elenca le virtù corporali, psicologiche e spirituali, oggetto di speculazioni filosofiche e teologiche lungo i secoli. L'elenco delle virtù da raggruppare è stato generato manualmente. Infatti l'interesse principale era di esplorare gruppi di virtù che costituiscono paradigmi tradizionali come fede, speranza e carità, rispetto ad altre virtù meno presenti in tavole tradizionali, come la salute fisica.

Come mostrato in Figura 7.6, i raggruppamenti sono chiaramente separati. Sul lato destro del grafico sono concentrate alcune virtù fisiche, arbitrariamente selezionate tra liste variabili di virtù fornite dai filosofi. Sul lato sinistro del grafico sono distribuite le sette virtù della tradizione cristiana, chiaramente distinte nei due gruppi consueti: le virtù teologali (fede, speranza e carità) nella parte alta del grafico e le virtù cardinali (forza, giustizia, temperanza e prudenza) nella parte inferiore.

<i>ἀρτιότης</i>	salute	<i>ἀνδρεία</i>	forza	<i>ἀγάπη</i>	carità, amore
<i>εὐαισθησία</i>	sensibilità	<i>δικαιοσύνη</i>	giustizia	<i>ἐλπίς</i>	speranza
<i>εὐεξία</i>	allenamento	<i>σωφροσύνη</i>	temperanza	<i>πίστις</i>	fede
<i>ὕγεια</i>	salute	<i>φρόνησις</i>	prudenza		

Tabella 7.7: Virtù

Altre categorie astratte testate con successo sono le emozioni, usando come seme le emozioni primarie tradizionali: *φόβος* (paura), *ἐλπίς* (speranza), *χαρά* (felicità) e *λύπη* (dolore).

Anche i raggruppamenti di termini astratti, come i gruppi di termini concreti visti sopra, possono mostrare la loro struttura gerarchica, espressa dalla presenza di sottoinsiemi significativi. Ad esempio, in molti esperimenti con diversi gruppi di virtù, le virtù civili *ἐλευθερία* (libertà) e *αὐτονομία* (autonomia) sono raggrup-

7. Esplorazione degli spazi semantici

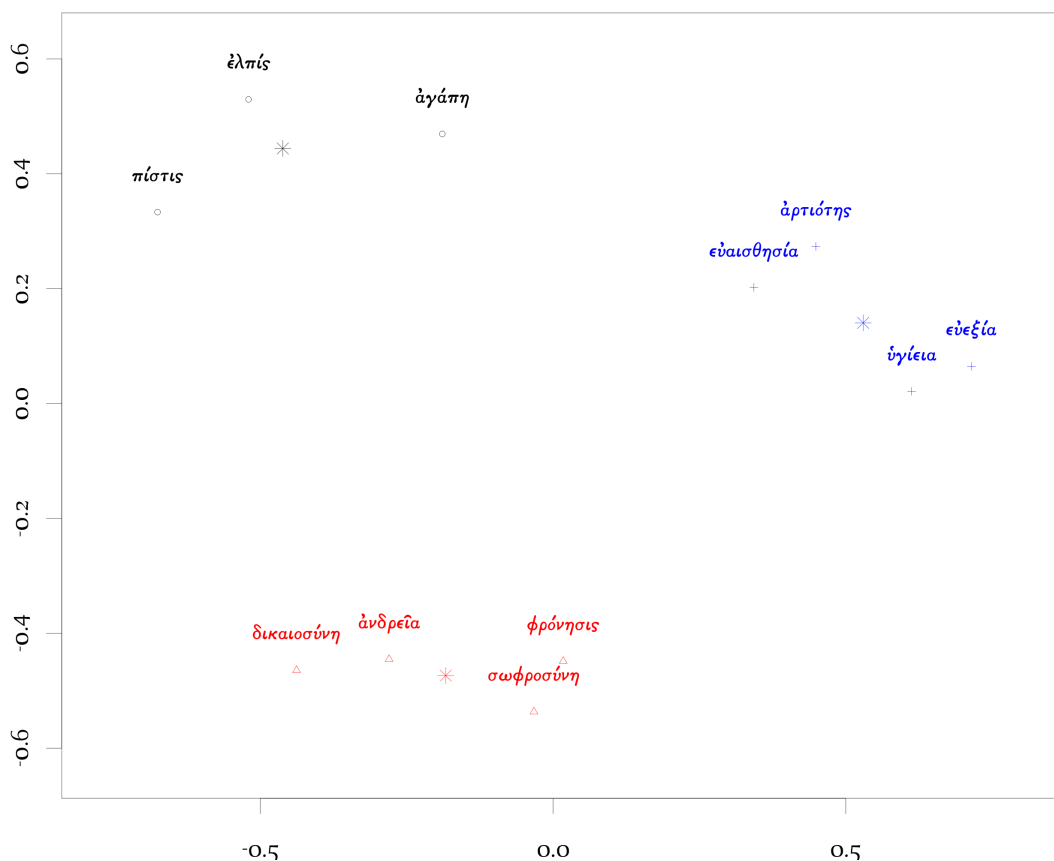


Figura 7.6: Raggruppamenti di virtù

pate in un sottoinsieme comune, a causa dell'implicazione logica di *αὐτονομία* e *ἐλευθερία*, in quanto l'una presuppone l'altra.

7.4 Spazi semantici e generi letterari

Rilevanti partizioni del *corpus* del TLG sono state esplorate per mostrare associazioni peculiari di tre generi letterari: genere epico, tragico e filosofico. Si sono scelte ancora una volta le parole *θάλασσα* (mare) e *θάνατος* (morte) per condurre l'indagine.

Poemi omerici

La dizione formulare dei poemi omerici facilita l'osservazione di associazioni coerenti all'interno dello spazio semantico, mostrando sia le relazioni sintagmatiche espresse dalle formule, come *πόδας ὠκὺς Ἀχιλλεύς* (Achille dal piede rapido), sia le relazioni paradigmatiche espresse dai sistemi formulari, come il nome proprio di uno stesso eroe seguito, in diverse parti dei poemi, da epiteti diversi. La coerenza semantica è osservabile sia per il termine concreto che per quello astratto. Nel caso di *θάλασσα* (27 occorrenze), le associazioni sono: *κῦμα* (onda), *κλύζω* (lavo, cado sopra: detto del mare), *κῆτος* (mostro marino), *θεμόω* (spingo la nave a terra), *ἔκβασις* (sbocco), *ἄλμυρός* (sale), *ἐπαγλαίζω* (onoro), *παλιρρόνιος* (refluente), *ἀελπής* (senza speranza), *περιμήκης* (molto grande). Per *θάνατος* (119 occorrenze) le associazioni principali sono: *τέλος* (fine), *μοῖρα* (destino), *πότμος* (destino avverso, morte), *ψυχή* (vita, anima) *κῆρ* (cuore), *ἐκπτύω* (sputo fuori l'ultimo respiro), *ὑπόβρυχα* (sott'acqua), *μόρος* (destino), *θνήσκω* (muoio), *φείδομαι* (risparmio persone o cose). Sia nel caso di termini concreti che in caso di termini astratti le associazioni sono significative o per ragioni paradigmatiche, come la sinonimia, o per ragioni sintagmatiche, come le scene tipiche.

Tragedia

A causa della bassa frequenza dei termini che caratterizzano la dizione tragica, non è possibile purtroppo apprezzare associazioni particolarmente significative.

Lo spazio semantico creato sulla partizione del TLG relativa alla Tragedia fornisce associazioni rilevanti solo per i termini che sono parole chiave nella dizione tragica, come *γός* (pianto) o *ἄτη* (rovina).

Le seguenti associazioni sono fornite dalle due parole usate per la prova. *θάλασσα* (34 ricorrenze) è associato a *ῥοιβδέω* (mi muovo con un suono sibilante), *σύνεγγυς* (vicino), *ἐρυμνός* (fortificato), *Ἐρυθραί* (Eritre), *συγκλύζω* (lavo), *Υψισται* (Tebe), *Θάας* (rapido: detto di navi), *πόντος* (mare aperto), *στέλεχος* (tronco), *τεράζω* (interpreto i prodigi). Migliore è la lista delle associazioni semantiche con il termine astratto *θάνατος* (111 occorrenze): *μέλεος* (miserabile), *δόλιος* (traditore),

7. Esplorazione degli spazi semantici

ἀνδρολέττειρα (assassino), *πάθος* (passione, sfortuna), *παναίτιος* (totalmente colpevole), *θεόκραντος* (compiuto dagli dei), *ισόψυχος* (di anima simile), *ἰχώρ* (sangue), *βαρύμηγυς* (pesante nell'ira), *φίλιος* (amichevole).

Mentre *θάλασσα* è associato solo a un piccolo gruppo di termini significativi e a una serie di altre parole apparentemente irrelate, *θάνατος*, parola chiave in Tragedia, fornisce associazioni più efficaci.

Filosofia

Lo spazio semantico della filosofia antica mostra la natura analitica delle associazioni in esso presenti: i termini concreti sono associati ai loro iperonimi, iponimi, co-iponimi, olonimi, meronimi e co-meronimi, oltre che a termini che esprimono attributi e qualità. I termini astratti sono principalmente associati ad antonimi e sinonimi o a parole che appartengono al *frame* della morte, come *γῆρας* (vecchiaia). Entrambi i termini sono associati anche a verbi semanticamente pertinenti. *θάλασσα* (1884 occorrenze) produce la lista di termini associati *ποταμός* (fiume), *ἄλμυρός* (sale), *πότιμος* (potabile), *λίμνη* (lago), *ἄλμυρότης* (sale), *πηγᾶς* (terra indurita dopo la pioggia), *πέλαγος* (mare aperto), *στάσιμος* (stagnante), *εἰσβάλλω* (dirigo verso il mare), *ῥέω* (fluire). *θάνατος* (2107 occorrenze) produce la lista *ἀποθνήσκω* (muoio), *ζωή* (vita), *θνήσκω* (muoio), *συμφορά* (disgrazia), *ζάω* (vivo), *διάλυσις* (dissoluzione), *χωρισμός* (separazione), *γῆρας* (vecchiaia), *τελευτάω* (termino la vita), *νεκρός* (cadavere).

I testi filosofici tendono a strutturare la conoscenza in tavole fisse, citate e discusse seguendo un ordine stabilito e dando a termini comuni significati tecnici. La tabella 7.8 elenca le antinomie pitagoriche, al fine di mostrare un esempio chiaro della differente distribuzione di speciali coppie antinomiche nello spazio semantico dell'intero TLG e nel *corpus* dei testi filosofici.

In Figura 7.7 si evidenziano con cerchi tratteggiati le tre classificazioni errate nello spazio semantico dell'intero *corpus* del TLG. I raggruppamenti corretti si incrociano; *καμπύλον* (curvo), è molto vicino a *πέρας* (limite); *ἑτερόμηκες* (rettangolo) è fortemente attratto dalla coppia *περιττόν* (dispari) – *ἄρτιον* (pari). Inoltre,

7.4 Spazi semantici e generi letterari

πέρας	limite	ἄπειρον	infinito
περιττόν	dispari	ἄρτιον	pari
ἓν	uno	πλῆθος	multitudine
δεξιόν	destra	ἀριστερόν	sinistra
ἄρρην	maschio	θῆλυ	femmina
ἡρεμοῦν	quiete	κινούμενον	movimento
εὐθύ	dritto	καμπύλον	curvo
φῶς	luce	σκότος	tenebra
ἀγαθόν	buono	κακόν	cattivo
τετράγωνον	quadrato	ἑτερόμηκες	rettangolo

Tabella 7.8: Antinomie pitagoriche

anche le coppie che sono correttamente classificate non mostrano una associazione forte.

Al contrario, lo spazio semantico del *corpus* filosofico mostra forti associazioni fra le coppie di antonimi pitagorici e anche le associazioni sbagliate sono meno gravi delle associazioni sbagliate viste nel grafico precedente relativo all'intero TLG. Nello spazio semantico filosofico i termini matematici *περιττόν* (dispari) e *ἄρτιον* (pari) sono chiaramente separati dai termini geometrici *τετράγωνον* (quadrato) e *ἑτερόμηκες* (rettangolo), mentre nello spazio totale del TLG si sovrappongono. Nello spazio semantico filosofico, *ἓν* (uno) e *πλῆθος* (moltitudine) sono correttamente isolati. I raggruppamenti errati relativi a *εὐθύ* (dritto), *καμπύλον* (curvo), *πέρας* (limite) e *ἄπειρον* (infinito) non si incrociano. Invece nello spazio del TLG i sei termini sono classificati male e sono vicini l'uno all'altro nella parte centrale del grafico.

7. Esplorazione degli spazi semantici

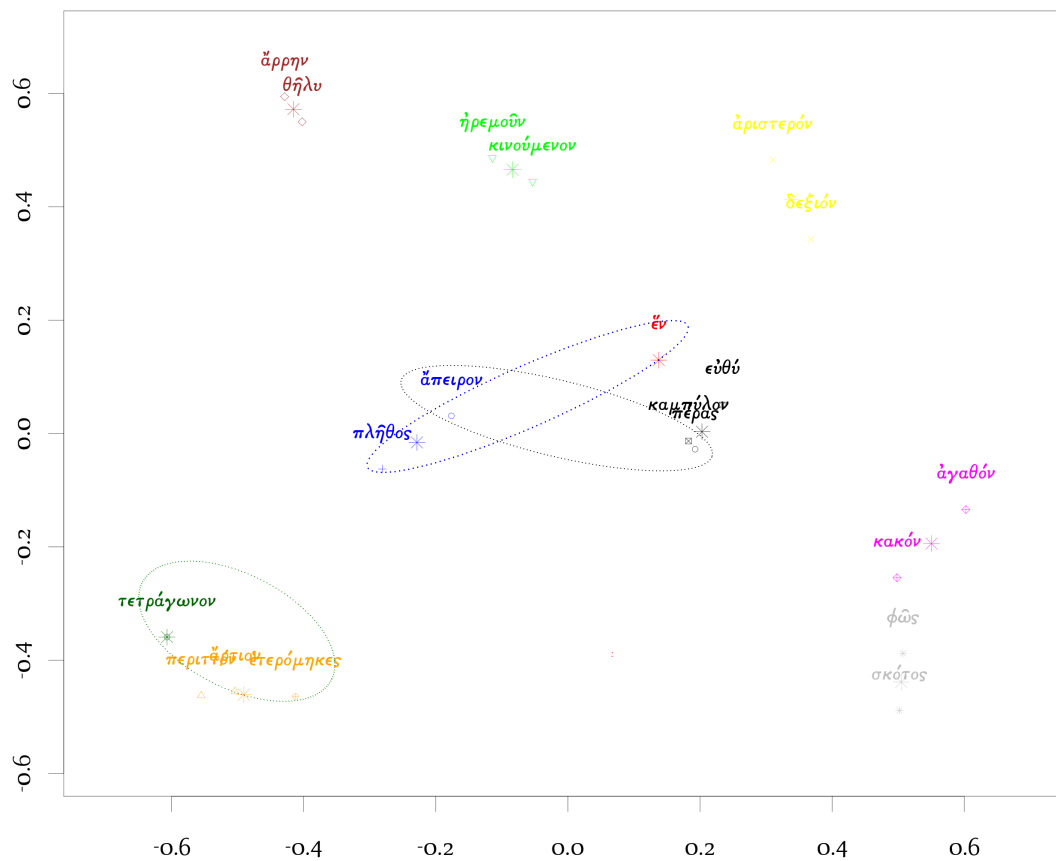


Figura 7.7: Coppie antinomiche nel TLG

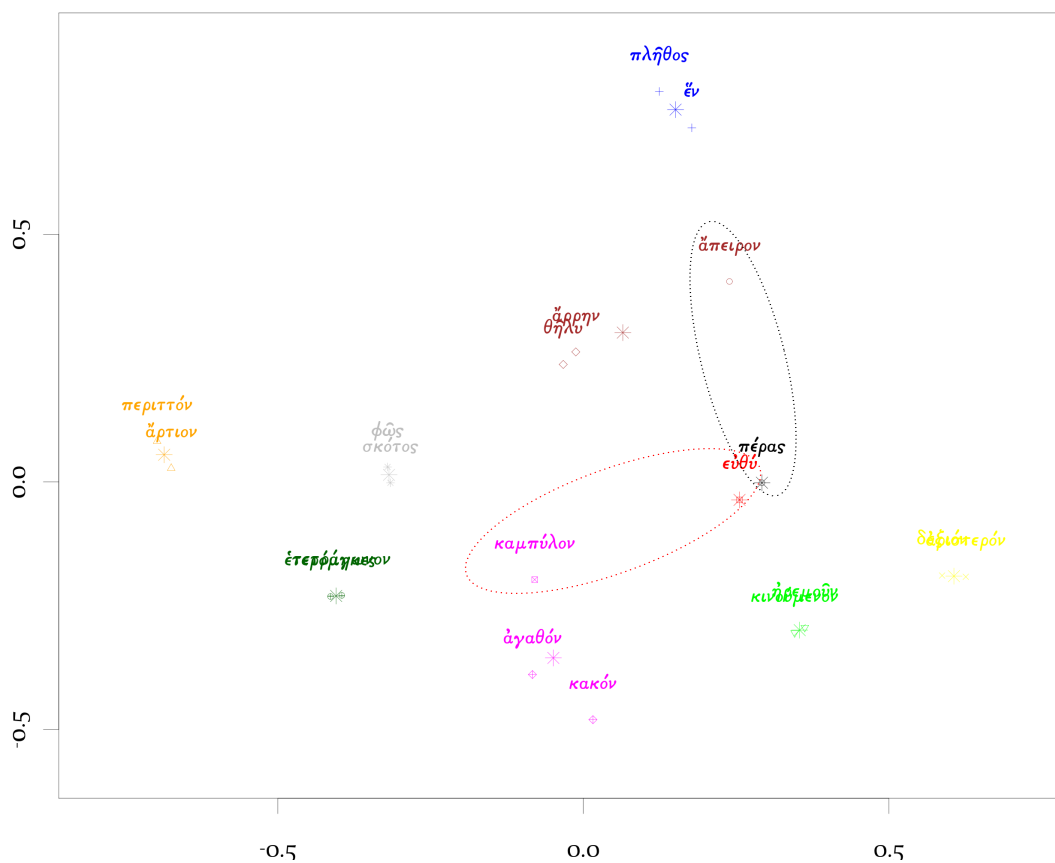


Figura 7.8: Struttura degli antonimi nel *corpus* dei filosofi

8

Commento a tre luoghi scelti

8.1 *Pers.* 114–139 e 532–547

ἀβροπενθείς: storia del problema

Pers. 133-139

λέκτρα δ' ἀνδρῶν πόθῳ
πίμπλαται δακρύμασιν
Περσίδες δ' ἀκροπενθείς ἐκά-
στα πόθῳ φιλάνορι
τὸν αἰχμήεντα θοῦρον εὐνα-
τῆρ' ἀποπεμψαμένα
λείπεται μονόζυξ.

La lezione di **M** ci presenta, senza equivoci, le giovani Persiane in preda a un dolore di forte intensità per il desiderio degli uomini che le hanno lasciate sole. Tuttavia, condividendo le ragioni espresse da PALEY (1871), la maggior parte delle edizioni moderne accoglie nel testo *ἀβροπενθείς* in luogo di *ἀκροπενθείς*, aprendo la strada a una molteplicità di interpretazioni sulla natura stessa delle donne Persiane. Anche in **Y** si legge *ἀβροπενθείς*, ma Paley non disponeva della collazione di questo codice e aveva formulato la sua congettura soltanto a partire

8. Commento a tre luoghi scelti

da considerazioni di natura paleografica, linguistica, stilistica e dal testo dello scolio¹ **M**:

ἐπιμόνως πενθοῦσαι, ὡς δοκεῖν ἀβρύνεσθαι ἐπὶ τῷ πενθεῖν

L'*ἀβρύνεσθαι* dello scolio sarebbe di commento a una differente lezione² rispetto a quella del testo di **M** e richiamerebbe quindi un *ἀβροπενθείς*, secondo Paley da accogliere senz'altro come genuino. Di parere nettamente opposto è MASSA POSITANO (1949, p. 63), la quale cita anche lo scolio **A**, derivante dallo scolio **M** (*αἱ δὲ Περσίδες γυναικες ἀβροπενθείς καὶ ἐπιμόνως πενθοῦσαι, ὡς δοκεῖν ἐναβρύνεσθαι ἐπὶ τῷ πενθεῖν*), per dire che

essi non sono altro se non una spiegazione dell'attrib. *ἀβροπενθείς*: gli scoliasti hanno inteso il componente *ἀκρο-* nel senso di *ἐπιμόνως* 'pervicacemente, continuamente', hanno cioè tradotto *ἀβροπενθείς* con 'continuamente piangenti' e hanno aggiunto alla traduzione il commento: *ὡς δοκεῖν ἀβρύνεσθαι ἐπὶ τῷ πενθεῖν*, 'si da sembrare di compiacersi del piangere'; *ἀβρύνεσθαι* è soltanto un commento, anzi, se si vuole, un fine commento, ma non altro, non certo un'implicita testimonianza dell'esistenza di una lezione *ἀβροπενθείς*.

Massa Positano, ignorando (come Paley) l'esistenza della lezione di **Y**, esordiva osservando che «tutti i codd. hanno nel v. 135 la lez. *ἀβροπενθείς*», e questa convinzione ha avuto quasi senza dubbio un'influenza determinante sul suo giudizio: «non si tratta qui della scelta fra due lezioni di autorità più o meno eguale: in realtà la lezione è una sola». Nella sua critica CITTI (1994, p. 34 n. 66) rileva che «*ἀβρύνεσθαι* introdurrebbe un'immagine del tutto inopportuna e impensabile se non illustrasse *ἀβροπενθείς* del testo», e così infatti pensa la maggior parte degli editori e degli studiosi dopo Paley. Voci fuori dal coro, ricordate da BELLONI (1994, *ad loc.*), restano solo GROENEBOOM (1930, *ad loc.*) e ITALIE (1964, *s.v.* *ἀβροπενθής*), i quali del resto non portano gli stessi argomenti di Massa Positano.

Lo scolio si pone quindi, per consenso quasi unanime, come *terminus ante quem* per datare la lezione *ἀβροπενθείς*, la quale verrebbe a essere così più antica

¹Per un approfondimento sul modo di procedere di Paley si veda RICHARDSON (1957).

²GARVIE (2009, *ad loc.*) osserva che lo scolio “shows an awareness of both readings: *ἐπιμόνως* is more appropriate to *ἀκρο-*, but *ἀβρύνεσθαι* does not look like a mere scholiast's comment”.

dei testimoni manoscritti che ci attestano la variante concorrente. **Y**, codice dell'inizio del XIV secolo, sarebbe uno dei tanti casi in cui un *recentior* reca un testo più antico rispetto al resto della tradizione.

Se i sostenitori di *ἀβροπενθείς* hanno messo molta enfasi sull'incompatibilità di *ἀβρύνεσθαι* con *ἀκροπενθείς*, CITTI (1994, *loc. cit.*) invita a considerare il fatto che la prima parte degli scoli **M** e **A** è a sua volta incompatibile con *ἀβροπενθείς*:

Se mai si può pensare che lo scolio [*scil.* sch. A] rifletta ambedue le varianti, dato che esso non solo contiene anche *ἀκροπενθείς*, ma lo glossa *καὶ ἐπιμόνως πενθοῦσαι* [e questo vale anche per il più antico sch. M], con un *interpretandum* che non funziona affatto per *ἀβροπενθείς*.

Prima che l'osservazione di Paley riscuotesse tanto favore, era addirittura lo scolio a essere corretto alla luce del testo tradito: nella sua edizione del 1885 Wecklein lascia nel testo *ἀκροπενθείς*, ma riporta in apparato, ritenendola degna di attenzione, «*ἀβροπενθείς* Paley»; poi nell'*Appendix* scrive «in scholio *ἐπιμόνως* legisse videtur», quasi che *ἐπιμόνως* (“continuamente, ininterrottamente”) non fosse adeguato a commentare il componente *ἀκρο-* della lezione accolta e occorresse postulare nello scolio un originario *ἐπιτόνως* (“molto intensamente”); il fatto che non si cogliesse la contraddizione fra le due parti dello scolio unite dal *καὶ* (meccanismo del resto non infrequente per giustapporre commenti a varianti) fa supporre che lo scolio venisse interpretato proprio nei termini proposti dalla Massa Positano e sentiti oggi tanto improbabili. Sulle interpretazioni dei termini in questione si tornerà in seguito.

Pur accettando l'ipotesi di Citti che gli scoli attestino entrambe le varianti (con un *καὶ* che avrebbe, di fatto, il valore di un *vel*, e anzi di un *aut*) è comunque impossibile dimostrare, solo attraverso gli scoli, quale sia la datazione relativa della parte che riguarda *ἀκρο-* e della parte che riguarda *ἀβρο-*. Se anche il solo *ἀβροπενθείς* risalisse ai primi scoli, è noto come varianti antiche non siano *tout court* garanzia di genuinità³.

Dal punto di vista paleografico, la probabilità che *ἀβροπενθείς* si sia corrotto in *ἀκροπενθείς* è pari alla probabilità che il secondo si sia corrotto nel primo. DAWE

³Basta sfogliare il capitolo *Varianti antiche e antiche edizioni* del libro di Pasquali per trovare numerosi esempi.

8. Commento a tre luoghi scelti

(1963, p. 130), che sostiene senza riserve la genuinità di ἀβροπενθείς, parlando di **Y**, unica testimonianza manoscritta, dice tuttavia che «writing β for κ is a very natural visual error in a pure minuscule script» e ritiene «perfectly possible» che la variante di **Y** sia un errore meccanico tale da riprodurre casualmente la lezione attestata dallo scolio, e secondo lui genuina. Accettare questa ipotesi, del tutto plausibile, significa però togliere autorità alla testimonianza di **Y** e dover tener conto soltanto degli scoli, i più antichi dei quali facevano riferimento a edizioni in maiuscola. Anche in questo caso Dawe, nella nota al passo citato, viene in soccorso:

Although the β, κ confusion is especially common in minuscule scripts, it is quite frequent at all times. Professor Page refers me to a very early example of this confusion at Sappho, Fr. 2, where αβροῖς (corrected to ἄβρωϛ by Bergk) is found in Athenaeus's citation, and ακρωϛ is on the ostrakon.

Tanto ἀβροπενθείς quanto ἀκροπενθείς sono ἄπαξ λεγόμενα, è necessario quindi vedere nel dettaglio quali argomenti giochino a favore dell'uno e dell'altro.

ἀβρός, ἀβροσύνη e altri derivati e composti: polisemie

In primo luogo sono necessarie alcune precisazioni su ἀβρός e i suoi derivati e composti, dato che la polisemia di questi termini può dar luogo a interpretazioni anche opposte fra di loro.

CHANTRAINE (1999), che ha difficoltà a fornire l'etimologia del termine⁴, dà come significato primario di ἀβρός «gracieux, délicat, joli»; tuttavia l'osservazione di SEMERANO (1994, *s.v.*), quand'anche non se ne accettassero le ragioni etimologiche di una derivazione dall'accadico ḥabāru, kabāru («to become fat, thick»), rende ragione in modo molto più efficace del rapporto fra significato primario e significati secondari: «il valore originario è “florido, soffice, paffuto, grasso”, e perciò “molle”». Fra queste due accezioni, anche se non è possibile stabilire con certezza quale sia primaria, si crea una polarità fra contrazione ed espansione⁵ che agisce sia a livello materiale che morale. La contrazione, la chiusura su di

⁴Pur segnalando una proposta di derivazione da ἥβη, si rende conto della sua improbabilità.

⁵Si veda FABBRI E MARRONE (2000, p. 244) per l'uso dei termini in questa accezione.

sé, catalizza i significati di /grazioso, delicato, molle, flebile, debole, indebolito/ sul piano fisico e di /effeminato, languido, rammollito/ sul piano morale, mentre l'espansione, l'apertura verso l'esterno, catalizza al contrario i significati di /lussureggiante, florido, vigoroso/ sul piano fisico e di /splendido, magnifico/ sul piano morale. Fra i due poli si colloca il significato di /raffinato, artefatto, sofisticato, lussuoso/, detto sul piano sia materiale che morale. Almeno negli autori prima di Platone, infatti, il concetto di raffinatezza espresso da *ἀβρός* e dai suoi composti non è associato costantemente e necessariamente al concetto di debolezza o di posizione recessiva; indica piuttosto un termine medio fra splendore ed effeminatezza.

La prima attestazione di *ἀβρός* si ha in Hes. fr. 339 M.-W., come epiteto di *παρθένος*. È poi largamente impiegato presso i lirici per caratterizzare le Grazie (Sapph., fr. 128 V.), Adone (fr. 140 V.), Andromaca (Sapph. fr. 44.7 V.) giovani fanciulle e giovani donne, bambini, parti del corpo come i piedi e la chioma. In una iscrizione è anche riferito a un piccolo animale da compagnia, considerato quasi un giocattolo: *ἀβρόν ἄθυρμα IG 14.1647 Lipara*.

In Pindaro si trova *ἀβρός* riferito a termini astratti come *πλοῦτος* (Pi. *P.* 3.110) oppure *κῦδος* (Pi. *O.* 5.7): si tenga conto che l'epiteto abituale di questo termine in Omero è *μέγα* (e.g. *Θ* 176, *Θ* 237, *I* 303, etc.) e che Esichio (*s.v.*) glossa *κῦδος* con *δόξα, ἀρετή, νίκη, ἀριστεία*. Niente a che fare quindi, in questo passo, con l'*ἀβροσύνη* intesa come mollezza incompatibile con l'esercizio della virtù. Anche nei composti che stanno a indicare termini morali come *ἀβρόβιος* (Bacchyl. 18.2 M.) o *ἀβροδίατος* (*Pers.* 41) la raffinatezza, la delicatezza di costumi non implica necessariamente un rilassamento, una mancanza di vigore: nei *Persiani* anzi si dice (vv. 41-43):

*ἀβροδιαίτων δ' ἔπεται Λυδῶν
ὄχλος, οἷτ' ἐπίπαν ἠπειρογενὲς
κατέχουσιν ἔθνος, [...]*

Non vi è incompatibilità quindi con l'esercizio del potere, in questo caso dei Lidi nei confronti dei popoli nati sul continente.

Al v. 543, a breve distanza da *ἀβρόγοι*, Eschilo usa *ἀβροχίτωνας* riferito ai letti delle giovani Persiane, mentre al v. 1073 Serse si rivolge al coro con l'esor-

8. Commento a tre luoghi scelti

tazione: γοᾶσθ' ἀβροβάται. BELLONI (1994) interpreta «mollemente incedendo», e cita Suda α 70 Adler Ἀβρὰ βαίνων. θρυπτόμενος. βλακεύμενος, apparentemente inoppugnabile, specie alla luce dell'interpretazione di KORZENIEWSKI (1967, p. 62): «der weichlich gazierte Schritt fällt ihm schwer, weil das persische Land durch Unglück δύσβατος [v. 1074] geworden ist». Anche se Eur. *Tro.* 819-20 μάταν ἄρ', ὧ χρυσέαις ἐν / οἰνοχόαις ἀβρὰ βαίνων, potrebbe essere di sostegno a questa linea interpretativa, per il fatto di trovarsi in un simile contesto di lamentazione, tuttavia due passi della *Medea* sembrano giustificare un esame più attento sulla polisemia di questo composto. Ai vv. 829-830 αἰεὶ διὰ λαμπροτάτου / βαίνοντες ἀβρῶς αἰθέρος è detto dei discendenti di Eretteo, chiamati, ai versi immediatamente precedenti (824-829):

Ἐρεχθείδαι τὸ παλαιὸν ὄλβιοι
καὶ θεῶν παῖδες μακάρων, ἱερᾶς
χώρας ἀπορθήτου τ' ἄπο, φερβόμενοι
κλεινοτάταν σοφίαν, [...]

“felici, figli degli dei beati, nutriti di sapienza”, tanto da essere a loro più confacente un passo leggiadro piuttosto di un indolente incedere. Ancora più chiaro *Med.* 1164 ἀβρὸν βαίνουσα, detto di Creusa che ha ricevuto in dono il peplo fatale. Ai vv. 1155-66 il suo entusiasmo e i suoi gesti sono descritti in modo puntuale: alla vista del dono è disposta a accettare tutte le condizioni del marito, poi prende il peplo, si mette la corona d'oro tra i riccioli e s'aggiusta la chioma allo specchio, sorride, poi ai vv. 1163-66:

κᾶπειτ' ἀναστᾶσ' ἐκ θρόνων διέρχεται
στέγας, ἀβρὸν βαίνουσα παλλεύκῳ ποδί,
δώροις ὑπερχαίρουσα, πολλὰ πολλάκις
τένοντ' ἐς ὀρθὸν ὄμμασι σκοπουμένη.

διέρχεται [...] ὑπερχαίρουσα, πολλὰ πολλάκις [...] sono poco compatibili con il βλακεύμενος («to be lazy», *LSJ*, s.v.) di Suda visto sopra. Il fatto che Creusa si guardi i talloni sollevati (τένοντ' ἐς ὀρθὸν) suggerisce che ἀβρὸν βαίνειν sia

un modo di muoversi in punta di piedi, quasi un passo di danza⁶. Così intende GROENEBOOM (1930, p. 232), mettendo in rilievo il contrasto fra l'incedere e il lamento: «men ziet, dat de op de spits van den voet zich bewegende dansers *ἀβροβάται* kunnen heeten, in een bittere tegenstelling met het voorafgaande *γοᾶσθε*». È curioso come questa accezione di *ἀβροβάτης* converga semanticamente verso *ἀκροβάτης*, termine attestato in epoca classica da iscrizioni di Efeso nel contesto del culto di Artemide⁷.

Non si intende certo proporre un emendamento, che non trova alcun supporto nella paradosi: interessa qui soltanto rilevare la compatibilità nei contesti presi in esame sia di composti con *ἀβρο-* sia di composti con *ἀκρο-*. Hermann congetture una seconda occorrenza di *ἀβροβάτης* al fr. 60 R., trasmesso dallo scolio ad Ar. *Av.* 276 come *ἀβρατεύς* ma, a ulteriore conferma delle facili confusioni fra *β* e *κ*, trasmesso come *ἀκρατοῦς* dal codice V.

In Tragedia, *ἀβρός* trova impiego riferito a fanciulle: Aesch. fr. 313 R. *παρθένους ἀβραῖς* o a parti del corpo: Soph. *Tr.* 523 *εὐώπις ἀβρά*, Eur. *Tro.* 506 e *Hel.* 1528 *ἀβρόν πόδα*, Ph. 1485-1486 *οὐ προκαλυπτομένα βοτρυώδεος ἀβρά παρῆδος* (col gen. dipendente dal neutro plurale). In questo passo delle *Fenicie* in effetti il termine può essere interpretato nel duplice senso di «delicato» oppure di «reso molle dal pianto», mentre l'associazione di delicatezza e debolezza è ancora più chiara in *IA* 614 *ἀβρόν τιθείσα κῶλον ἀσθενές θ' ἄμα*.

In Erodoto *ἀβρός* occorre in 1.71 *Πέρσησι γάρ, πρὶν Λυδοὺς καταστρέψασθαι, ἦν οὔτε ἀβρόν οὔτε ἀγαθὸν οὐδέν*, dopo il celebre discorso di Sandanis per dissuadere Creso dalla spedizione contro i Persiani, dove le raffinatezze dei Lidi non possedute dai Persiani si riducono a un piccolo elenco: vesti in tessuto al posto di abiti in pelle e disponibilità di vino e di fichi. Ma l'aspetto più importante non è tanto la contrapposizione fra beni materiali diversi, quanto piuttosto l'opposizione fra scelta e necessità: i Persiani, al contrario dei Lidi, *σιτέονται δὲ οὐκ ὅσα ἐθέλουσι, ἀλλ' ὅσα ἔχουσι, χώρην ἔχοντες τρηχέα*. In questo, con l'opposizione di *ἀβρός*

⁶Lo Pseudo-Zonaras, *s.v.* *ἀβραβαίνων* scrive: *θρυπτόμενος κολακεύομενος, ἐκείνος ὄντως ἀβραβαίνων ἐδόκει ἔχων τὰς ὀφρῦς ὑπερηγμένους ἄνω*. Un modo di camminare, quindi, proprio di chi si compiace di se stesso.

⁷Si veda *LSJ*, *s.v.*

8. Commento a tre luoghi scelti

a *τραχύς*, si accorcia la distanza fra il significato di raffinato e quello di florido. Diverso è il caso della seconda occorrenza erodotea: 4.104 Ἀγάθυρσοι δὲ ἀβρότατοι ἄνδρες εἰσὶ καὶ χρυσοφόροι τὰ μάλιστα, dove si ritrova l'associazione ἀβροσύνη - χρυσός dei *Persiani* di Eschilo. Come secondo elemento di composto compare in *Λυδὲ ποδαβρέ*, contenuto in un oracolo della Pizia (1.55).

Si è visto che ἀβρὸν πόδα e ἀβρὸν βαίνειν trovano una continuità semantica nel composto ἀβροβάτης, così l' ἀβρὸν βόστρυχον di Eur. *Bacch.* 493 ha un corrispettivo nel composto ἀβροκόμης, riferito metaforicamente alla chioma degli alberi da Euripide in *IT* 1099 e *Ion* 920 φοίνικα [...] ἀβροκόμαν, ma anche alla chioma umana al di fuori della Tragedia, in Orph. *H.* 56.2 e in numerosi altri luoghi, fra cui *AP* 12.256, con un piacevole gioco fra i due significati. *IT* 1149 ἀβροπλούτιο χαίτας indica la capigliatura riccamente lussureggiante, mentre in Omero, *P* 439, la chioma è detta θαλερή. In Omero è da notare anche l'epiteto ἀκρόκομοι (*Δ* 533) riferito ai Traci, con il significato di «summo vertice comantes⁸»; significativi sono gli impieghi di questo composto da parte di Nonno di Panopoli: nelle *Dionisiache* infatti compare sia ἀβροκόμης che ἀκρόκομος. Il composto con ἀβρο- è usato in una decina di casi con riferimento a persone (10.426 ἀβρ-ης *Ληνεύς* 13.91 ἀβρ-ης Ὑμέναιος 13.456 ἀβρ-ων Ἐρώτων 13.559 ἀβρ-οιο κασιγνήτιο 16.172 ἀβρ-ην ἀκοίτην 33.81 ἀβρ-ης Ὑμέναιος 42.126 ἀβρ-ω κούρη 42.243 ἀβρ-ου Ἐνδυμίωνος 48.148 ἀβρ-ου κούρης 48.356 ἀβρ-ω *Κυθερείη*) e in un luogo soltanto, per altro fortemente corrotto, denota le alture ricoperte di vegetazione: 15.47 ἀβρ-ων κορύμβων. Al contrario il composto con ἀκρο-, nelle sue tre occorrenze, non è mai riferito a persone: 11.502 ἀκρ-οισι κορύμβοις 15.112 ἀκρ-ου φοίνικος 12.56 ἀκρ-ους φοίνικας. Si vede quindi che lo stesso tipo di albero detto ἀβροκόμης in Euripide si trova a essere ἀκρόκομος in Nonno, mentre all'interno dello stesso poema i κόρυμβα possono essere caratterizzati ora dall'uno, ora dall'altro epiteto.

In Pindaro (*P.* 8.89 ἀβρότατος ἔπι) il termine astratto indica la condizione della freschezza giovanile, mentre in Euripide (*Bacch.* 968-970) si assiste a un gioco fra ἀβρότης e τρυφή. L'interpretazione forse è soggettiva, ma sembra di scorgere una gradazione di significato fra il semplice farsi portare e il farsi portare ἐν χειρί.

⁸Si veda EBELING (1885, *s.v.*).

ΔΙ. Φερόμενος ἤξεις ... ΠΕ. Ἀβρότητ' ἐμὴν λέγεις.
 ΔΙ. Ἐν χερσὶ μητρός. ΠΕ. Καὶ τρυφᾶν μ' ἀναγκάσεις.
 ΔΙ. Τρυφᾶς γε τοιάσδ'. ΠΕ. Ἀξίων μὲν ἄπτομαι.

L'astratto ἀβροσύνη di Or. 349 indica infine il lusso di Menelao.

Un discorso a parte merita il composto ἀβρόπηνος, che fa riferimento alla ricchezza degli abiti: Lyc. 863 ἀβροπήνους [...] πέπλους. In Ag. 690 ἀβροπήνων προκαλυμμάτων è congettura del Salmasius per ἀβροτίμων della paradosi. Fraenkel la considera una «palmary emendation», dato che ritiene la lezione tradita una corruzione facile: a sostegno di questa ipotesi cita Paley, il quale osserva come in PV 223 τιμαῖς al posto di ποιναῖς sia offerta da P^{γρ}. Tuttavia l'esempio non è felice: il passaggio da -πήνων a -τίμων troverebbe giustificazione solo sul piano del significante, cioè per ragioni paleografiche, mentre ποιναῖς e τιμαῖς sono in relazione anche sul piano del significato. Inoltre si può pensare che il verso della corruzione sia opposto rispetto a quello ipotizzato da Fraenkel, in quanto τιμαῖς nell'accezione di “punizioni” è *difficilior* rispetto a ποιναῖς⁹. GRIFFITH (1984, *ad loc.*) ritiene quest'ultima una glossa intrusiva che banalizza, appiattendo sull'interpretazione univoca di “pene da patire”, l'ambiguità carica di sarcasmo di τιμαῖς, che significa allo stesso tempo «'honours' (i.e. rewards), and 'retribution' (i.e. punishment)», ottenuti da Prometeo in cambio dei consigli dati a Zeus durante la Titanomachia.

A una conclusione simile era giunto anche Friis Johansen, parlando di una congettura di Blaydes a *Suppl.* 24 che proponeva di emendare βαρύτιμοι con βαρύποιοι:

the corruption of -ποιοι to -τιμοι is (despite Ag. 690 ἀβροπήνων Salmasius: ἀβροτίμων codd., kept by Page) improbable, the word ποιινή being sometimes used to explain τιμή in its pejorative sense [...] it occurs as an inferior variant [...] and may plausibly be regarded as such in PV 223

Le considerazioni di Paley viste sopra sembrano quindi non costituire un valido sostegno alla congettura del Salmasius: più convincente invece, semmai, la

⁹Si veda DAWE (1963, p. 146). JUDET DE LA COMBE (1982, II, p. 33) rileva giustamente come l'apparato di Page non attribuisca τιμαῖς soltanto a una nota marginale di un codice ma ad almeno altri due testimoni della tradizione manoscritta (in West la *varia lectio* è attribuita alla famiglia δ, costituita dai codici ADW).

8. Commento a tre luoghi scelti

presenza di *ἀβροπήνων* proprio in Licofrone, un poeta che notoriamente prende in prestito dal lessico eschileo¹⁰. Tuttavia Fraenkel stesso invita alla prudenza: «It is probable that Lycophron 863 borrowed *ἀβροπήνους ... πέπλους* from Tragedy; we cannot be sure whether he took it from Aeschylus».

Judet de La Combe osserva la superiorità espressiva del tradito *ἀβροτίμων*, che accoglie nel testo, rispetto al congetturale *ἀβροπήνων*, accettato fra gli altri da Page e da West: la lezione salmasiana restituisce «le détail d'une "scène de genre"», attraverso un epiteto che nulla aggiunge alle qualità attese dai *προκαλυμμάτων* di Elena mentre, secondo la sua interpretazione, la lezione della paradosi ha la virtù

d'établir un lien entre la qualité du lieu quitté, qui vaut par son prix, et Troie, où est chanté l'hymne "en l'honneur de l'épousée" [...]»

Per finire questo rapido *excursus*, rimane da prendere in considerazione il derivato verbale *ἀβρύνω*. L'occorrenza eschilea di Ag. 919 mette in relazione il termine sia con i modi femminili, sia con gli eccessi orientali (vv. 918-920):

*Καὶ τ᾿ἄλλα μὴ γυναικὸς ἐν τρόποις ἐμὲ
ἄβρυνε, μηδὲ βαρβάρου φωτὸς δίκην
χαιμαίετες βόαμα προχάνης ἐμοί*

Al medio (Aesch. *Ag.* 1205, Soph. *OC* 1339, Eur. *IA* 858) assume invece il significato di «superbire¹¹», compiacersi, darsi delle arie, derivato dal senso originario di vivere in modo raffinato.

Per LOMBARDO (1983, p. 1079), che dedica un ampio studio al significato del termine in epoca arcaica (ma solo rapidi cenni all'epoca classica), il mutamento semantico di *ἀβρός* sarebbe avvenuto dopo le guerre persiane:

[...] ciò che conferisce ad *habrosyne* la sua specificità e il suo peculiare interesse consiste nel fatto che, a differenza di altre nozioni, come ad esempio *tryphé* e *chlidé*, essa è attestata già in epoca arcaica. Inoltre, sia il termine in questione che l'aggettivo *habrós* appaiono connotati in un primo tempo in chiave positiva [...] e solo successivamente e non uniformemente in chiave negativa; un atteggiamento critico verso le *habrosynai anophelees* è già

¹⁰Si veda JUDET DE LA COMBE (1982, I, p. 34).

¹¹Si veda ITALIE (1964, *s.v.*).

in Senofane (fr. 3 G.-P.), ma è solo dopo le guerre persiane e in ambiente greco-continentale che la connotazione cambia in maniera più decisa, contemporaneamente all'emergere documentario di altre nozioni consimili come *chlidé* e poi *tryphé*.

A sostegno della sua tesi dice in nota:

Cf. in particolare A., *Pers.*, 41; 135; 541 sq.; 1073; A., 690; 918 sq.; e poi anche Pl., *Alc. I*, 122b; X., *Cyr.*, VIII, 8, 15 sq.; Antiphan., fr. 91 Kock

Tuttavia a noi sembra che l'analisi dei contesti, in particolare per quanto riguarda i *Persiani*, faccia trasparire piuttosto una situazione di incertezza semantica, dove la connotazione morale negativa sta forse emergendo e soppiantando i valori positivi del termine, ma non ha ancora completamente neutralizzato l'antica accezione di “florido, rigoglioso”, etc.

È bene comunque approfondire lo studio di altri contesti e in particolare delle cooccorrenze con i termini semanticamente affini. Il concetto di *ἀβροσύνη*, come rileva Lombardo, è molto spesso associato, dall'età classica in poi, al concetto di *τρυφή*¹². Benché, come si vedrà fra breve, i lessicografi antichi stabilissero per lo più fra i due termini un rapporto di stretta sinonimia, tuttavia è necessario fare alcune distinzioni.

Etimologicamente, *τρυφή* è un sostantivo derivato per dissimilazione dell'aspirazione dalla radice verbale **θρυφ-*, da cui proviene anche il verbo *θρύπτω* che significa «broyer, briser, ramollir», tanto che il composto *ἐνθρύπτω* si dice di pane lasciato immerso in un liquido (Hp. *Salubr.* 7 ἄρτος ἐν οἴνω ἐντεθρυμμένος). In Omero *τρύφος* (δ 508) ha il significato di «fragmentum, pars refracta». Tanto la serie dei derivati in *θρυπ-* quanto la serie in *τρυφ-* ha sviluppato dei significati metaforici sul piano sia materiale che morale: *θρυπτικός*, con significato attivo, indica la capacità di spezzare le pietre in Dsc. 1.121, mentre con significato passivo e traslato indica effeminatezza in X. *Cyr.* 8.8.15; *θρύψις* significa “rompere in piccoli pezzi” in Arist. *GC* 316b30; indica invece la mollezza del corpo in Plu.

¹²La ricerca sul TLG dimostra che su circa un migliaio di occorrenze di termini in *ἀβρο-*, distribuite lungo tutto l'arco della grecità, più del dieci per cento si trovano in associazione più o meno diretta con termini contenenti *τρυφ-* o *θρυπ-*.

8. Commento a tre luoghi scelti

Dem. 4 e la mollezza dello spirito in *X. Cyr.* 8.8.16; *τρυφερός* indica la delicatezza di parti del corpo in *Batr.* 66, ma la voluttà in *Ar. V.* 551, e così via¹³.

Eschilo presenta *Ag.* 1595 *ἔθρουπτ'* col significato proprio di sminuzzare, detto di Tieste che fa a pezzi le carni dei propri figli, e *PV* 891 *τῶν πλούτῳ διαθρυπτομένων*, col significato traslato di «luxuriare¹⁴». Nessuna occorrenza invece di *τρυφ-*, che non viene evidentemente mai associato da Eschilo alla «mollezza» orientale. Oltre ad alcune testimonianze nei frammenti dei Tragici minori, fra cui *e.g.* *Lyc.* 607 e il frammento sofocleo 275.818 Radt, i termini della famiglia di *τρυφή* sono prediletti da Euripide: *τρυφή* fr. 54.2, 892.4, 36.2 Schnell, *Supp.* 552, *Tro.* 997, *Ph.* 1491, *Or.* 1113, *Bacch.* 970; *τρυφάω* *Supp.* 214, *Ion* 1376, *Bacch.* 969, *IA* 1303; *ἐντρυφάω* fr. 362.24, fr. 53.24 De Gruyter, *Cycl.* 588; *τρυφερός* *Bacch.* 150; *τρυφήμα* *IA* 1050.

Degno di nota è *IA* 1303-4:

ἄ μὲν ἐπὶ πόθῳ τρυφῶσα
Κύπρις, ἄ δὲ δορὶ Παλλάς

in quanto esprime un'idea simile a quanto espresso dall'*ἐναβρύνεσθαι* dei nostri scoli eschilei.

La correlazione fra le due serie di termini e la contiguità semantica è testimoniata già in epoca classica: Eur. *Ph.* 1486-1491 raffigura Antigone in lacrime e si sofferma sulla guancia e sulla veste con due enallagi molto efficaci: *ἀβρὰ παρηῆδος* [...] *στολίδος κροκόεσσαν* [...] *τρυφάν*. Di nuovo, in *Bacch.* 968-70, i due termini sono a stretto contatto: *ἀβρότητ' ἐμὴν λέγεις*. / [...] *καὶ τρυφᾶν μ' ἀναγκάσεις / ΔΙ. τρυφάς γε τοιάσδ'*. Nell'ambito della Commedia, è l'*ἀβροσύνη* degli Ioni a essere caratterizzata da Antiph. 91 con le parole: *Ἰώνων / τρυφεραμπεχόνων ἀβρὸς ἡδυπαθῆς / ὄχλος*.

L'unica occorrenza platonica di *ἀβρός* e le due occorrenze di *ἀβρότης* sono fortemente rivelatrici del modo in cui può essersi fissata l'equivalenza semantica con *τρυφή*.

In *Smr.* 204c *καὶ γὰρ ἔστι τὸ ἐραστὸν τὸ τῷ ὄντι καλὸν καὶ ἀβρὸν καὶ τέλεον καὶ μακαριστόν* è ancora in vigore la mappa dei valori dell'età arcaica.

¹³Si veda CHANTRAINE (1999, *s.v.*), per una esemplificazione più esauriente.

¹⁴Si veda ITALIE (1964, *s.v.*).

Prima, in *Smpr.* 197d Eros è detto *τρυφής, ἀβρότητος, χλιδῆς, χαρίτων, ἰμέρου, πόθου πατήρ ἐπιμελής ἀγαθῶν*. La serie intende quindi connotare una sequenza di termini positivi, alla quale segue la sequenza delle circostanze negative nelle quali Eros è guida e ottimo salvatore: *ἀμελής κακῶν ἐν πόνῳ, ἐν φόβῳ, ἐν πόθῳ, ἐν λόγῳ*¹⁵ *κυβερνήτης, ἐπιβάτης, παραστάτης τε καὶ σωτῆρ ἄριστος*. In *Alc.* 122c *ἀβρότης*, di nuovo in associazione con *τρυφή*, è questa volta oggetto di biasimo esplicito:

εἰ δ' αὖ ἐθέλεις εἰς πλούτους ἀποβλέψαι καὶ τρυφᾶς καὶ ἐσθῆτας ἱματίων θ' ἔλξεις καὶ μύρων ἀλοιφᾶς καὶ θεραπόντων πλήθους ἀκολουθίας τῆν τε ἄλλην ἀβρότητα τῆν Περσῶν, αἰσχυθείης ἂν ἐπὶ σεαυτῷ, αἰσθόμενος ὅσον αὐτῶν ἐλλείπεις.

Al di fuori dell'associazione con *ἀβρός* e *ἀβρότης* il concetto di *τρυφή* è per lo più impiegato in sequenze di termini che indicano debolezza, mollezza, etc.: *Grg.* 492c *τρυφή καὶ ἀκολασία καὶ ἐλευθερία*, 525a *ὑπὸ ἐξουσίας καὶ τρυφής καὶ ὕβρεως καὶ ἀκρατίας*, *R.* 422a *τρυφήν καὶ ἀργίαν καὶ νεωτερισμὸν*, 556b *ἀρ' οὐ τρυφῶντας μὲν τοὺς νέους καὶ ἀπόνους καὶ πρὸς τὰ τοῦ σώματος καὶ πρὸς τὰ τῆς ψυχῆς, μαλακοὺς δὲ καρτερεῖν πρὸς ἡδονάς τε καὶ λύπας καὶ ἀργούς*; 590b *τρυφή δὲ καὶ μαλθακία*, *Ti.* 65d *τραχύτερα μὲν ὄντα τρυφνά*, *Criti.* 121a *ἀλλ' οὐ μεθύοντες ὑπὸ τρυφής διὰ πλοῦτον ἀκράτορες αὐτῶν ὄντες ἐσφάλλοντο*, *Lg.* 637e *Πέρσαι δὲ σφόδρα μὲν χρώνται καὶ ταῖς ἄλλαις τρυφαῖς ἄς ὑμεῖς ἀποβάλλετε, ἐν τάξει δὲ μᾶλλον τούτων*, 695b *παραλαβόντες δ' οὖν οἱ παῖδες τελευτήσαντος Κύρου τρυφής μεστοὶ καὶ ἀνεπιπληξίας*, 695c *Δαρεῖος γὰρ βασιλέως οὐκ ἦν ὕος, παιδεία τε οὐ διατρυφώση τεθραμμένος*, 695d *μετὰ δὲ Δαρεῖον ὁ τῆ βασιλικῆ καὶ τρυφώση πάλιν παιδευθεὶς παιδεία Ξέρξης*, 806c *τέλεον γὰρ καὶ οὐ διήμισυν δεῖν τὸν νομοθέτην εἶναι, τὸ θῆλυ μὲν ἀφιέντα τρυφᾶν καὶ ἀναλίσκειν διαίταις ἀτάκτως χρώμενον, τοῦ δὲ ἄρρενος ἐπιμεληθέντα, τελέως σχεδὸν εὐδαίμονος ἥμισυ βίου καταλείπειν ἀντὶ διπλασίου τῆ πόλει*, 900e *ἀμέλειαν τε καὶ ἀργίαν καὶ τρυφήν εἰς ἀρετὴν ψυχῆς θήσομεν, ἢ πῶς λέγεις*; 901a *τρυφῶν καὶ ἀμελής ἀργός τε* 901c *ῥαθυμία καὶ τρυφή*, 901e *ῥαθυμία δὲ ἀργίας καὶ τρυφής*.

Negli scoli e nei lessici l'identificazione dei due termini è ormai stabile. L'*Etymologicum Magnum*, s.v. "Ὀβρυζος" scrive: *παρὰ τὸ ἀβρόν, ὃ σημαίνει τὸ τρυφερόν*.

¹⁵Da notare la ripetizione polisemica di *πόθος*, da un lato a chiudere un'anticlimax di termini che esprimono condizioni gradevoli, dall'altro al centro di un'anticlimax che esprime situazioni di disagio (specie se si intende *λόγος* come "disputa").

8. Commento a tre luoghi scelti

πένθος - *γόος*: *il situational script omerico (corrispondenze con i nostri passi)*

Opportunamente SIDERAS (1971, p. 212) individua nei vv. 133s. *λέκτρα δ' ἀνδρῶν πόθῳ / πίμπλαται δακρύμασιν* l'eco di una locuzione omerica: Hom. *υ* 58 *κλαίει δ' ἄρ' ἐν λέκτροισι καθεζομένη μαλακοῖσιν*, in riferimento a Penelope. Del resto *λέκτροισι ... μαλακοῖσιν* trova un altro richiamo nel più elaborato *λέκτρων εὐνάς ἀβροχίτωνας* del v. 543. Si cercherà di vedere se anche la caratterizzazione omerica del *πένθος* e del *γόος* possa dare luce ai due passi della parodo e del primo stasimo che stiamo prendendo in esame.

Nei poemi epici *πένθος* copre i campi semantici di *luctus* e di *maeror*. In quanto affezione dell'anima, si manifesta *ἐνὶ στήθεσσι* (*P* 139), *ἐνὶ φρεσὶ* (*Σ* 88, *η* 217, *σ* 324, *ω* 423), *μετὰ φρεσίν* (*Ω* 105), *ἐν κραδίῳ* (*ρ* 489), affligge l'animo nell'intimo: *X* 242 *ἔνδοθι θυμὸς ἐτείρετο πένθει λυγρῷ* e può estendersi all'intera persona: *Π* 548 *κατὰ κρήθεν λάβε πένθος*, fino a diventare esso stesso una condizione all'interno della quale ci si trova, come dice Andromaca fra i singhiozzi: *X* 483 *αὐτὰρ ἐμὲ στυγερῷ ἐνὶ πένθει λείπεις*. In generale, esso agisce come una forza fuori controllo e viene descritto da Omero al suo sopraggiungere: *ικάνει* (*ζ* 169) / *ἵκετο* (*A* 362, *Σ* 64, *Ψ* 224, *Ω* 708) / *καθίκετο* (*α* 342), al suo aumentare: *ἀέξει* (*λ* 195) / *ἄεξε* (*ρ* 489) / *ἀέξων* (*P* 139, *ω* 231,) oppure quando ha preso posizione in una sede dell'anima: *ἔκειτο* (*ω* 423) o è posseduto stabilmente da qualcuno: *ἔχουσα* (*Ω* 105), *ἔχοντα* (*η* 217, *κ* 376). Per contro, non viene mai ripreso nella fase egressiva. Questo fatto è del resto in linea con gli epiteti di *πένθος*, che, oltre a indicarne l'aspetto affettivo, ne determinano soprattutto la dimensione e l'intensità, attraverso un articolato sistema formulare piuttosto coerente: da un lato quindi *στυγερῷ* (*X* 483), *λυγρῷ* (*β* 70, *X* 242), *χαλεπὸν* (*ζ* 169), *ἀρητὸν* (*P* 37, *Ω* 741), *ἄλαστον* (*Ω* 105, *α* 342, *ω* 423) e dall'altro *μέγα*¹⁶ (*A* 254, *Δ* 417, *H* 124, *P* 139, *λ* 195, *ρ* 489), *κρατερόν* (*Δ* 249, *κ* 376), *μυρίον* (*Σ* 88), *ἀμέτρητον* (*τ* 512), *ἀάσχετον* (*Ω* 708), *ἀτλήτω* (*I* 3).

Il verbo derivato *πενθέω* è impiegato due volte, su un totale di quattro occorrenze, all'interno di una formula che esprime l'idea di quanto sia negativo

¹⁶Per mezzo di questo epiteto è costruito anche il nome proprio *Μεγαπένθης*, figlio di Menelao e di una schiava (*ζ* 10, 11, *ο* 100, 103, 122, 141).

affliggersi senza fine: *ἐπεὶ κάκιον πενθήμεναι ἄκριτον αἰεὶ* (in σ 174 detto da Eurinome a Penelope e in τ 120 detto da Odisseo come considerazione generale). Anche gli aggettivi derivati si concentrano sull'aspetto dell'intensità dei dolori, sulla loro molteplicità oppure sulla capacità della persona a cui è attribuito l'epiteto di tollerarli; si hanno così *νηπενθές* (δ 221), detto del *φάρμακον* dell'oblio impiegato da Elena, che esprime il grado più basso della scala; *πολυπενθής* (I 563, ξ 386, ψ 15), che indica sia l'intensità sia la molteplicità dei dolori, provati rispettivamente dall'alcione, da Odisseo e dal cuore di Penelope; *ταλαπενθέα* (ϵ 222), riferito al *θυμός* di Odisseo. Si distingue dagli altri il composto *νεοπενθής*, usato una sola volta in Omero ma ripreso in epoca tarda, soprattutto da Nonno di Panopoli, che lo impiega nelle *Dionisiache* ben quattro volte (8.286, 19.1, 37.100, 11.119). Nell'Odissea è riferito alle giovani inesperte della vita, contrapposte in particolare ai vecchi che hanno molto sofferto (λ 38-9):

*νύμφαι τ' ἡίθεοί τε πολύτλητοί τε γέροντες
παρθενικάί τ' ἀταλαὶ νεοπενθέα θυμὸν ἔχουσαι*

In Omero il *γόος* invece indica prevalentemente l'espressione del dolore attraverso un'emissione sonora di varia intensità e variamente articolata. Quando Afrodite cerca conforto dalla madre, Dione le prospetta i singhiozzi di Egialea, la compagna di Diomede, se l'eroe dovesse trovare la morte (*E* 412-414):

*μῆ δὴν Αἰγιάλεια, περίφρων Ἀδρηστίνη,
ἐξ ὕπνου γοώουσα φίλους οἰκῆας ἐγείρη,
κουρίδιον ποθέουσα πόσιν, τὸν ἄριστον Ἀχαιῶν*

Viene introdotta così la figura della donna che si profonde in alti lamenti, da sola, nella notte, rimpiangendo lo sposo che si è macchiato di *ὑβρις*. La somiglianza con i nostri passi dei *Persiani* tuttavia si ferma qui, dato che in Eschilo il modulo è moltiplicato per un numero indefinito di volte ed esteso a tutta la città, dove ciascuna donna rimane anonima e piange un uomo altrettanto privo di nome.

Scelta ancora più felice, se si considera che il *γόος* si può esprimere sì nella dimensione personale (e in questa forma caratterizza la figura di Penelope), ma trova il suo naturale sviluppo nella dimensione pubblica, dove può assumere

8. Commento a tre luoghi scelti

il senso generico di lamento oppure il significato specifico del lamento funebre, omaggio per il morto (*Ψ* 9, *ω* 19 *ὁ γὰρ γέρας ἐστὶ θανόντων*)¹⁷ e fortemente codificato, tanto da essere iniziato, ma sarebbe meglio dire intonato, da un membro di spicco della collettività, maschile o femminile, e seguito da tutti gli altri, solitamente raggruppati anch'essi secondo il genere. Si viene ad avere quindi lo schema formulare: *τοῖσι/τῆσι(ν)* [...] + Nom. + [...] (*ἐξ*)*ἦρχε γόοιο*, che vede protagonisti rispettivamente *Θέτις* (*Σ* 51), il *Πηλεΐδης* (*Σ* 316, *Ψ* 17), *Ἑκάβη* (*X* 430, *Ω* 747), *Ἀνδρομάχη* (*Ω* 723) e per finire *Ἑλένη* (*Ω* 761), che intona il lamento *τριτάτη*, dopo Andromaca ed Ecuba.

Come viene iniziato volontariamente, così il *γόος* può essere fatto cessare sotto il controllo della volontà, se almeno così si può interpretare l'affermazione di Menelao *δ* 102-3 *ἄλλοτε μὲν τε γόω φρένα τέρπομαι, ἄλλοτε δ' αὐτε / παύομαι*. Inoltre si può chiedere di smettere il pianto, pur senza averne rimosso la causa, ma per pure ragioni di opportunità: ad esempio per prestare ascolto, come in *τ* 268, oppure per non farsi sentire, come in *φ* 228. Naturalmente l'arresto del pianto può essere inteso anche come rimozione della sua causa, ad esempio in *P* 38, dove il figlio di Pantoo dice che, se fosse in grado di uccidere Menelao, *γόου κατάπαυμα γενοίμην*, oppure in *δ* 800, dove Atena manda il fantasma di Iftime affinché faccia cessare (*παύσειε*) il lamento di Penelope lenendo l'afflizione del suo animo attraverso l'annuncio del ritorno di Telemaco.

A differenza del *πένθος*, che sopraggiunge come una forza agente e autonoma, il *γόος*, o meglio il desiderio del *γόος*, viene spesso indotto da una persona a un'altra persona oppure a una collettività: con il verso formulare "*Ως φάτο, τῷ δ' ἄρα πατρὸς ὑφ' ἡμερον ὄρσε γόοιο* Priamo spinge Achille al lamento in *Ω* 507 mentre la stessa formula è riferita a Menelao nei confronti di Telemaco in *δ* 113. In *τ* 249 e in *ψ* 231, invece, il verso formulare "*Ως φάτο, τῆ/τῷ δ' ἔτι μᾶλλον ὑφ' ἡμερον ὄρσε γόοιο* chiude la mozione degli affetti rispettivamente di Odisseo verso Penelope e di questa verso lo sposo appena riconosciuto. Speculare è anche l'abbraccio fra Telemaco e il padre, in *π* 215 *ἀμφοτέροισι δὲ τοῖσιν ὑφ' ἡμερος ὄρτο γόοιο*, mentre

¹⁷In *Ψ* 9 in realtà si dice: *Πάτροκλον κλαίωμεν ὁ γὰρ γέρας κτλ.*, tuttavia il v. 10 continua con *αὐτὰρ ἐπεὶ κ' ὄλοοιο τεταρπόμεσθα γόοιο*, rendendo lecito l'accostamento a *ω* 19.

lo stesso Telemaco in *ρ* 46 chiede alla madre di non indurlo al lamento: *μητρ ἐμή, μή μοι γόον ὄρνυθι*. La formula *ὕφ' ἡμερος ὄρσε γόοιο* è impiegata poi, riferita a collettività, in *Ψ* 14, *Ψ* 108, 153, *δ* 183 e il medesimo concetto viene espresso in *Z* 499, *Ω* 760 e *κ* 398, anche se non vengono impiegate formule specifiche. In *κ* 248 *γόον δ' ὠίετο θυμός* è invece l'animo stesso il soggetto della brama.

Il *γόος* del resto è desiderato (*κ* 398 *ἡμερόεις ὑπέδν γόος*) perché produce una forma di piacere simile a quella di altre necessità fisiche, come ad esempio mangiare e bere. Non è casuale quindi il fatto che *Ω* 227 *ἐπὴν γόου ἐξ ἔρον εἶην* sia costruito sullo schema del frequente *αὐτὰρ ἐπεὶ πόσιος καὶ ἐδητύος ἐξ ἔρον ἔντο*, con 21 occorrenze, la prima delle quali in *A* 269. Le espressioni più frequenti per esprimere questa idea sono costruite con il verbo *τέρπω*: *Ψ* 98, *λ* 212 *τεταρπώμεσθα γόοιο* – *δ* 102 *ἄλλοτε μὲν τε γόω φρένα τέρπομαι* – *τ* 213, 251, *φ* 57 *ἢ δ' ἐπεὶ οὖν τάρφθη πολυδακρύτοιο γόοιο* – *ψ* 10 *αὐτὰρ ἐπεὶ κ' ὄλοοιο τεταρπώμεσθα γόοιο* – *Ω* 513 *αὐτὰρ ἐπεὶ ῥα γόοιο τετάρπετο δῖος Ἀχιλλεύς*. È rilevante inoltre il fatto che venga ripreso così spesso il momento in cui si cessa dal lamento per passare ad altre attività. Infatti, come per i desideri del ventre, anche per il *γόος* si raggiunge presto la sazietà: *δ* 103 *αἰψήρως δὲ κόρος κρυεροῖο γόοιο*, e inoltre: *Ψ* 157 *γόοιο μὲν ἔστι καὶ ἄσαι*.

Per quanto voluto e cercato, gli epiteti ne mostrano il profondo legame con le circostanze di dolore nel quale si sviluppa, creando degli apparenti ossimori rispetto ai verbi di desiderio impiegati; tali epiteti sono: *ἀρητὸν*¹⁸ (*P* 37, *Ω* 741), *ὄλοοιο* (*Ψ* 10, 98), *κρυεροῖο* (*Ω* 524, *δ* 103, *λ* 212), *ἀλίσστον* (*Ω* 760), *οἰζυροῖο* (*θ* 540), oppure *δακρυόεντος* (*δ* 801, *ρ* 8, *ω* 323), *πολυδακρύτοιο* (*τ* 213, 251, *φ* 57). Del resto, la concomitanza di pianto e lamento è talmente naturale da potersi fondere in una sola immagine, attraverso una efficace metonimia: *δ* 758 *ὡς φάτο, τῆς δ' εὔνησε γόον, σχέθε δ' ὄσσε γόοιο*. Due epiteti si discostano dagli altri per caratterizzare l'intensità del *γόος*: *ἀδινού* (*Σ* 316, *X* 430, *Ψ* 17, *Ω* 747)/*ἀδινόν* (*δ* 721) e *θαλερόν* (*κ* 457).

¹⁸Sul significato di *ἀρητός* si veda GIORDANO (1998), dove viene ricostruito il senso di «lamento che maledice, che chiama vendetta». Il confronto con *Cho.* 315-331, dove si parla di *γόος ἔνδικος* (v. 330) attesterebbe l'esistenza della prassi di invocare la vendetta nel lamento funebre per un morto ucciso.

8. Commento a tre luoghi scelti

Lo scolio a Ψ 17 spiega $\acute{\alpha}\delta\iota\nu\omicron\upsilon$ γόοιο parlando del modo in cui viene emesso il fiato per produrre il lamento: $\delta\epsilon\acute{\iota}$ γὰρ ἐν θρήνοις προΐεναι μὲν ἄθροον τὸ πνεῦμα, μὴ δυνάμεον δὲ φέρειν ὑπεκλύεσθαι. Per quanto riguarda il secondo epiteto, invece, Eust. vol. 1, p. 388, l. 35 lo interpreta in questo modo: $\Theta\alpha\lambda\epsilon\rho\acute{o}s$ δὲ γόος καθ' ὁμοιότητα τοῦ $\theta\alpha\lambda\epsilon\rho\acute{o}\nu$ δάκρυ τὸ ἄπαστον καὶ αἰὲν νεάζον. E aggiunge, come attestano anche gli scoli: Ἀριστοφάνης δέ, φασι, στυγερὸν γράφει. $\Sigma\tau\upsilon\gamma\epsilon\rho\acute{o}s$ del resto è tramandato dalla tradizione indiretta come *varia lectio* rispetto a $\kappa\rho\upsilon\epsilon\rho\omicron\iota\omicron$ di δ 103, rispetto al quale però sembra rappresentare piuttosto una semplice glossa. In realtà i campi semantici di $\acute{\alpha}\delta\iota\nu\omicron\varsigma$ e di $\theta\alpha\lambda\epsilon\rho\acute{o}s$ possono avere parziali sovrapposizioni, se si considera che la nozione di «fitto» può avere un risvolto sia locale che temporale¹⁹ e che l'idea di «florido, rigoglioso», può suggerire l'idea di addensamento, per esempio quando è riferito alle chiome. Tuttavia, la distinzione semantica dei due epiteti, l'uno riferito quindi alla modalità di esecuzione e l'altro alla continuità temporale, torna comoda soprattutto a chi preferisca eliminare l'equivalenza delle due espressioni formulari isometriche²⁰.

In tutti i contesti presi in esame, si è visto che il motivo del $\pi\acute{\epsilon}\nu\theta\omicron\varsigma$ e il motivo del γόος possono essere correlati ma che mantengono caratteristiche ben distinte. La spontaneità del $\pi\acute{\epsilon}\nu\theta\omicron\varsigma$ opposta alla possibilità di inibirne le manifestazioni esteriori è ben rappresentata in ρ 489-91:

*Τηλέμαχος δ' ἐν μὲν κραδίη μέγα πένθος ἄεξε
βλημένου, οὐδ' ἄρα δάκρυ χαμαὶ βάλεν ἐκ βλεφάρουϊν,
ἀλλ' ἀκέων κίνησε κάρη, κακὰ βυσσοδομύων.*

Ma il prototipo del *situational script*²¹ omerico che regola il rapporto fra il dolore infinito che proviene dall'esterno e il lamento volontario che ha funzione autoconsolatoria si trova ai versi τ 512-517:

*αὐτὰρ ἐμοὶ καὶ πένθος ἀμέτρητον πόρε δαίμων
ἦματα μὲν γὰρ τέρπομ' ὀδυρομένη, γοώωσα,
...
αὐτὰρ ἐπεὶ νύξ ἔλθη ...*

¹⁹Così interpreta CHANTRAINE (1999).

²⁰Ovviamente $\acute{\alpha}\delta\iota\nu\omicron\varsigma$ e $\theta\alpha\lambda\epsilon\rho\acute{o}\nu$ sono isometrici nel contesto di κ 457, non in tutti i contesti possibili, in quanto il primo inizia per vocale e il secondo per consonante.

²¹Per il termine, si veda MINCHIN (2001).

Gli scoli non percepiscono questa opposizione, e trattano *πένθος* e *γόςος* come sinonimi, tanto che l'uno può essere impiegato per spiegare l'altro: sch. Ω 227 <ἐπήν γόου ἐξ ἔρον εἶην:> ἐπὰν κορεσθείην τῆς τοοῦ πένθους ἐπιθυμίας. Una volta stabilita l'equivalenza semantica fra un termine che indica in origine un'affezione dell'animo e un termine che indica in origine un'espressione esteriore di tale affezione, non risulta difficile proporre di emendare *οπε ingenii* Ω 584 μὴ ὁ μὲν ἀχθυμένη περ κραδίη χόλον οὐκ ἐρύσαιτο come ci indica lo scolio: Ἡρωδιανὸς γόον γράφει ποῖον γὰρ οὗτος εἶχε χόλον, εἰ μὴ μᾶλλον γόον;

L'interpretazione nelle traduzioni (in funzione attributiva oppure predicativa)

Prima che l'osservazione di Paley ricevesse il consenso quasi unanime delle generazioni successive di filologi, il rapporto fra il v. 135 e il v. 540 era di segno esattamente opposto, in quanto alla luce del concetto espresso nella parodo si rilevavano difficoltà interpretative o guasti da sanare nel primo stasimo. Così, mentre *ἀκροπενθεῖς* sembrava esprimere un concetto del tutto accettabile e tradizionale, *ἀβρόγοι* costituiva, almeno a un primo esame, una controversa impertinenza rispetto al contesto. Giocava a sfavore inoltre la vicinanza a soli due versi di *ἀβροχίτωνας*, che urtava la sensibilità estetica dei filologi a causa della ripetizione del primo elemento del composto.

Le «molliter gementes Persides», che nella parodo della traduzione di Stanley erano «impense tristes», creano problema già a Pauw, che nel proprio commentario osserva:

Nonne scribendum *ἀκρόγοι Περσίδες?* *ἀκρόγοος*, ut *ἀκρόχολος* et alia gemella: τὸ ἀβρόν non convenit hic in γόοις: et sequitur *ἀβροχίτωνας* suo jure: attende etiam ad *ἀκορετάτοις* sequens.

A questi risponde HEATH (1762), con diversa sensibilità:

ἀκρόγοι suum sibi habeat Pauwius. Novae nuptae sponsos recens amissos plorantes eleganter et accommodatissime *ἀβρόγοι* dicuntur. Talium enim lamentis tenerum quoddam ac delicatum, vigente scilicet adhuc et effervescente amore, inesse solet, quod in senum ac matronarum luctu personarum decoro non adeo convenit. Neque tamen illarum citius quam harum

8. Commento a tre luoghi scelti

lacrymas exarescere ideo existimandum est. Contra, illarum γόοι aptissime dicuntur ἀκορέστατοι, illas enim magis urere solet tam cari capitis desiderium.

Il giudizio su questo luogo continua a essere oscillante, con proposte di emendare ora il primo ora il secondo elemento del composto. Così BLOMFIELD (1823, *ad loc.*):

Vocem ἀβρόγοι, quam exhibent libri, obelo notavit Porsonus; nec stare potest, quum brevi intervallo sequatur ἀβροχίτωνες. [...] In simillimo loco v. 140 Περσίδες dicuntur ἀκροπενθείς. [...] In hoc versu olim conjeceram αἱ λαβρόγοι. Quidnam in suo exemplari invenerit vetustior e Scholiastis, conjectura assequi nequeo, nisi forte αἱ δακνόγοι, explicat enim, αἱ τοῖς γόοις δακνόμενοι [sic!].

A questo scolio, che non è riportato in alcuna edizione moderna, fa riferimento anche BOTHE (1831, *ad loc.*), dal quale ricava una diversa congettura, non registrata nel repertorio di Wecklein:

insulsum ἀβρογόοι, cum statim easdem mulieres dolere dicat γόοις ἀκορεστοτάτοις. Schol.: καὶ αἱ τοῖς γόοις δακνόμεναι, quasi legisset καὶ δηξίγοοι: verum hoc et inusitatum est, et in eandem reprehensionem incurrit.

Hartung accoglie nel testo della propria edizione la congettura καὶ δακρύροοι, con la nota: «Besserung nach dem Schol.»²². Il testo dello scolio recenziere in effetti contiene πολύκλαυτοι, che apparentemente non può spiegare ἀβρογόοι. Il ragionamento come si vede è simile a quello di Paley, ma orientato in maniera diversa.

Gli emendamenti al primo elemento del composto, come si è visto, sono tutti volti a intensificare l'espressione del dolore. Degno di nota è λαβρόγοοι, dove l'aggettivo impiegato abitualmente per indicare l'irruenza dell'acqua, del vento o del fuoco determina l'impeto del γόος attraverso un'efficace metafora naturalistica.

²²È da rilevare che la congettura non è menzionata nei repertori e Wecklein attribuisce ad Hartung καὶ δακρύροοι. Del resto, si è visto sopra come sia naturale e largamente attestata l'associazione di lacrime e γόος, di lacrime e πένθος, tanto da dare origine ad alcune formule. Zakas propone addirittura di emendare ἀκροπενθείς del v. 140 con δακρυπενθείς, preferendo un accostamento di concetti più tradizionale rispetto a composizioni audaci.

Del resto la plausibilità paleografica, se si pensa a un errore di maiuscola che vede la sequenza *ΔΑΒΡ*- ridursi a *ΛΑΒΡ*-, conferisce alla congettura un certo fascino, per quanto priva di ogni altro sostegno²³.

Gli emendamenti al secondo elemento del composto, invece, sono volti a descrivere qualità fisiche o morali delle donne persiane, attraverso l'idea tradizionale della delicatezza femminile, come *ἀβρόγυιοι* di Bothe, che adduce giustificazioni di carattere paleografico, o *ἀβροχρόοι* di Schütz, oppure attraverso l'idea emergente in età classica della mollezza connaturata ai costumi orientali, come *ἀβρόγονοι* di Lobeck o *ἀβρόβιοι* di de Marcellus. Hermann non rileva difficoltà in questo passo, e dichiara: «Praeter necessitatem hoc vocabulum coniecturis tentatum est». La mutata sensibilità della critica doveva spostarsi infatti dal primo stasimo alla parodo.

Dalle traduzioni si vede chiaramente che *ἀκροπενθείς* era inteso a caratterizzare le donne persiane nella circostanza contingente del loro *πόθος*: «impense tristes» per Stanley, «épouses désolées» per La Porte du Theil, «summa tristitia oppressa» per Bothe, «summo dolore oppressae» per Ahrens, etc. Più articolata invece la rosa di interpretazioni del restaurato *ἀβροπενθείς*. Paley si mantiene vicino alla linea esegetica dello scolio e traduce: «luxurious even in their grief», lasciando nella lingua moderna l'ambiguità fra sovrabbondanza, eccesso da un lato, raffinatezza, sofisticazione da un altro lato e mollezza da un altro ancora. Il concetto espresso nel primo stasimo è sentito invece in maniera diversa, e tradotto con «daintily sighing». Ma l'idea di tenerezza e di mollezza non tarda a estendere il suo dominio per tutto il secolo scorso: «tenderly mourning» dice Sidgwick, e ritiene che «in 541 we have the exactly similar word», che spiega nella sua Introduzione dicendo «in luxury of sorrow», e aggiungendo subito: «i.e. unchecked, weak, abandonment to grief. *ἀβρός* is 'soft', 'delicate', 'luxurious'». Per Smyth ciascuna di queste donne si trova «in the tenderness of her grief», per Mazon «en son deuil languissant», per Belloni le donne «si struggono languendo» e Judet de La Combe parla delle «femmes perses au deuil délicat».

²³Questa congettura e le congetture in *δάκρ*-, eliminando il nesso *αί δ(έ)*, tolgono la necessità di ripartire le donne in diversi gruppi.

8. Commento a tre luoghi scelti

Le corrispondenze fra 114-139 e 532-547

Si è detto fin dall'inizio come i passi 115-139 e 532-547 siano in stretta relazione tematica, in quanto ciò che viene temuto nella parodo viene descritto nel suo compimento nel primo stasimo. Vi è così una sapiente tessitura di corrispondenze semantiche su schemi metrici (prevalenza di leccio e cretico / sequenza di anapesti) e strutture sintattiche (inanimato in funzione di soggetto / soggetto animato) diverse.

Come fa notare DE ROMILLY (1958, p. 42), che segnala numerosi *loci paralleli*,

Tous les exemples que nous avons passé en revue supposent une action de la crainte dans le corps. Elle se manifeste dans le cœur, le diaphragme, le foie, les entrailles.

Così, per esprimere l'idea che il coro teme di ricevere l'informazione che l'esercito è perduto, la costruzione sintattica assume una struttura di grande efficacia stilistica. Il soggetto logico si trova nel caso obliquo (*μοι*), il soggetto sintattico è la *φρήν* rivestita metaforicamente a lutto (*μελαγχίτων*); il *verbum timendi* è espresso da un *nomen actionis* che regge la completiva, il cui soggetto ancora una volta non è il coro, ma un suo olonimo, cioè la città nel suo complesso. *Τούδε*, retto da *πύθηται*, è epanalettico rispetto al genitivo esclamativo²⁴ (vv. 114-119):

ταῦτά μοι μελαγχίτων
φρήν ἀμύσσειται φόβῳ
“ὄα, Περσικοῦ στρατεύματος,”
τούδε μὴ πόλις πύθη-
ται, κένανδρον μέγ' ἄστυ Σουσίδος

L'antistrofe non è meno sorprendente dal punto di vista sintattico (vv. 120-125):

²⁴MATINO (1998, p. 55), che cita questo passo, scrive:

Il genitivo di esclamazione spiega la fonte, la ragione dell'azione della principale e può quindi intendersi come un genitivo di causa. È accompagnato da una interiezione esclamativa.

καὶ τὸ Κισσίων πόλισμ'
 ἀντίδουπον ἄσεται,
 “ὄᾶ”,
 τοῦτ' ἔπος γυναικοπλη-
 θῆς ὄμιλος ἀπύων,
 βυσσίνους δ' ἐν πέπλοις πέσῃ λακίς.

ἄσεται è congettura di Burney per il tradito ἔσσεται di **B^aV**+λ oppure ἔσεται del resto della tradizione, metricamente impossibile. Considerando l'equivalenza metrica con ἔσσεται, «la correction ἄσεται n'enlève rien à l'impossibilité syntaxique», secondo Roussel. Belloni al contrario considera la frase al futuro una parentetica, ma accoglie la congettura di Burney. Mettendo insieme le due considerazioni, ci sembra che si possa mantenere il tradito ἔσσεται, copula di ἀντίδουπον, rendendo invece τοῦτ' ἔπος epanalettico rispetto a ὄᾶ e avendo in questo modo una maggiore simmetria rispetto alla strofe. In ogni caso, anche accettando la congettura di Burney, si verrebbero ad avere due completeive rette da φόβῳ del v. 115 interrotte da una parentetica al futuro e da un *nominativus pendens*, che tuttavia costituiscono il soggetto logico della seconda completeiva, che ha per soggetto sintattico il *nomen actionis* λακίς, in linea con le particolarità sintattiche della strofe. Del resto, nulla aggiunge dal punto di vista semantico ἄσεται, dato che il sema /sonoro/ si trova già in ἀντίδουπον.

La strofe successiva distende il tono su un paragone dal sapore omerico, fra l'esercito di terra e uno sciame di api mentre l'antistrofe si apre focalizzando un oggetto inanimato, togliendo dalla scena il soggetto logico, cioè le donne, rappresentate dal complemento di abbondanza δακρύμασιν (vv. 131-132):

λέκτρα δ' ἀνδρῶν πόθῳ
 πῖμπλαται δακρύμασιν

Solo a quel punto soggetto logico e soggetto sintattico vengono a coincidere, dopo la messa a fuoco, realizzata tramite un pregnante *nominativus pendens*, dal generale al particolare (vv. 135–139):

Περσίδες δ' ἀβροπενθεῖς ἐκά-
 στα πόθῳ φιλόνορι

8. Commento a tre luoghi scelti

τὸν αἰχμᾶεντα θοῦρον εὐνα-
τῆρ' ἀποπεμφαμένα
λείπεται μονόζυξ.

Rispetto a questo quadro, i versi 532-547 ripristinano i soggetti animati. Il soggetto del primo periodo sintattico è Zeus e la sintassi segue con rigorosa coerenza la struttura determinante + determinato, sia nel genitivo che precede il sostantivo, sia nel participio congiunto che precede la forma del verbo finito, sia nell'accusativo che precede il verbo (vv. 532-36):

ὦ Ζεῦ βασιλεῦ, νῦν <δὴ> Περσῶν
τῶν μεγαλᾶρχων καὶ πολυάνδρων
στρατιᾶν ὀλέσας
ἄστυ τὸ Σούσων ἠδ' Ἀγβατάνων
πένθει δνοφερῶ κατέκρυψας.

Le corrispondenze sul piano lessicale e semantico sono abbastanza chiare:

Περσῶν ... στρατιᾶν : *Περσικοῦ στρατεύματος*

ἄστυ τὸ Σούσων : *ἄστυ Σουσίδος*

mentre le corrispondenze semantiche sono più sottili:

gli attributi della città sono cambiati di segno e riferiti ai Persiani: se la città è *κέανδρον*, i Persiani²⁵ sono *πολυάνδρων* e mentre quella è *μέγ(α)*, questi sono *μεγαλᾶρχοι*.

Nell'antistrofe la costruzione della parodo diventa diretta: non è la lacerazione a cadere sulle vesti, ma sono le donne a farsi protagoniste (vv. 537-540):

πολλὰ δ' ἀπαλαῖς χερσὶ καλύπτρας
κατερεικόμεναι
διαμυδαλέους δάκρυσιν κόλπους
τέγγουσ' ἄλγους μετέχουσαι.

²⁵Per enallage, dato che l'epiteto dovrebbe essere di *στρατιᾶν*, anche non accettando, in quanto non sufficientemente motivata, la congettura di Wecklein al v. 9 per il tradito *πολυχρύσου στρατιᾶς*.

Se nella parodo erano i pepli a ricevere l'epiteto, qui *καλύπτρας* è senza epiteto, trasferito invece alla sfera dell'umano, nelle *ἀπαλαῖς χερσὶ*. Là i letti si riempivano di lacrime, senza nemmeno nominare l'agente, qui invece la stessa scena è trasferita con forza sulla persona, bagnando di lacrime un petto già umido (vv. 543–545).

λέκτρων εὐνὰς ἀβροχίτωνας,
 χλιδανῆς ἤβης τέρψιν, ἀφείσαι,
 πενθοῦσι γόοις ἀκορεστοτάτοις

Ora *λέκτρων* è divenuto addirittura genitivo di specificazione del complemento oggetto, in questi versi dove ritroviamo anche il *πένθος* e dove è ripresa l'idea di insaziabilità, di continuità che abbiamo visto anche ai versi precedenti e che richiama il *πίμπλαται* della parodo. Le *εὐνὰς ἀβροχίτωνας* / *χλιδανῆς ἤβης τέρψιν*, vengono abbandonate, e resta spazio soltanto al dolore insaziabile.

Conclusioni

Il quadro che ci si è delineato ci porta a dire che non ci si dovrebbe limitare a interpretare *ἀβροπενθείς* in base ad *ἀβρογοοί* o in relazione agli altri termini composti con *ἀβρο-* presenti nella tragedia. Se è vero che il termine ha modificato il suo significato dopo le guerre persiane, tuttavia solo i contesti non ambigui possono accertare se il cambiamento è stato rapido e netto, oppure lento e graduale. Abbiamo cercato di dimostrare che i contesti della nostra tragedia possono lasciare aperta la possibilità che *ἀβρός*, fra i molteplici significati, conservi ancora quello di «florido, lussureggiante, abbondante», come risultato di una metafora vegetale, e che *ἀβροπενθείς* non significhi altro che le donne *πενθοῦσι γόοις ἀκορεστοτάτοις*.

Del resto, se si studiano i composti di *ἀκρο-*, ci si accorge che i contesti in cui sono impiegati lasciano aperta un'altra ambiguità, in quanto *ἀκρο-* conferisce generalmente il significato di “superficialmente” e non di “estremamente”, anche se lo *Stephanus*, il *LSJ* e il *Diccionario Griego-Español*, sempre molto attento ai problemi di semantica, differiscono nelle interpretazioni. In fondo, questa più che la conclusione era il punto di partenza di Paley, nel suo commento *ad loc.*:

8. Commento a tre luoghi scelti

The MSS. and edd. give ἀκροπενθεῖς. This ought to mean, 'grieving from the depths of the heart,' but in Ag. 778 and Eur. Hec. 242, οἶδ', οὐ γὰρ ἄκρας καρδίας ἔφρασε μου, the sense is clearly the reverse, 'the mere surface of the heart.' So Hippol. 255, μὴ πρὸς ἄκρον μύελον ψυχῆς. Yet in Bacch. 203, δι' ἄκρων φρενῶν has the same force as τοξότης ἄκρος, Ag. 611, viz. that of height and superiority. Blomfield quotes ὀργὴν ἄκρος, Herod. i. 73, which contains the same idea. I have admitted the correction which I formerly proposed without noticing at the time that the Schol. must have so read, ὡς δοκεῖν ἀβρύνεσθαι ἐπὶ τὸ (λ. τῷ) πενθεῖν. Compare αἱ ἀβρόγιοι Περσίδες inf. 543. On β and κ confused see Suppl. 541.

Così, per concludere, forse lo scolio, accostando ἐπιτόμως e ἀβρύνεσθαι ha colto dal contesto la polisemia del solo ἀβροπενθεῖς e non ha espresso invece il significato di due varianti differenti.

8.2 Pers. 598–602

I vv. 598-602 sono resi da M in questo modo:

φίλοι, κακῶν μὲν ὅστις ἔμπορος κυρεῖ,
ἐπίσταται βροτοῖσιν ὡς ὅταν κλύδων
κακῶν ἐπέλθῃ, πάντα δειμαίνειν φιλεῖ,
ὅταν δ' ὁ δαίμων εὐροῇ, πεποιθέναι
τὸν αὐτὸν αἰεὶ δαίμων' οὐριεῖν τύχης.

Rispetto a M, West accoglie ἔμπειρος dei *recentiores* in luogo di ἔμπορος e la congettura di Weil αἰὲν ἄνεμον in luogo di αἰεὶ δαίμων'. Considerando anche che i due interventi testuali non sono dettati da ragioni metriche, sembra opportuno verificare il loro carattere di necessità, che Belloni stesso non vede ed è portato a difendere la paradosi con solidi argomenti.

In realtà ἔμπειρος sembra *lectio facillior* rispetto a ἔμπορος, almeno se si presta fede a Roussel:

Il est impossible qu'un scribe, trouvant le mot ἔμπειρος, qui va bien, et qui est courant dans sa langue, ait corrigé en ἔμπορος, mot qui, pour lui ne signifie que: marchand! ἔμπορος est ici avec un sens étymologique: celui qui n'est pas ἄπορος, avec un ἐν- positif, opposé au ἀ- négatif (ἄγαμος, ἔγγαμος et mille autres). H. Estienne (s. v.) qualifie ἔμπορος de ἐμφατικώτερον καὶ ποιητικώτερον, c'est qui est excellemment dit.

Eppure, molti editori hanno sollevato riserve su ἔμπορος di natura logica e semantico-sintattica, e la preferenza per ἔμπειρος risale già a Robortello. Il dubbio sulla consequenzialità logica della sentenza è bene espresso da HEATH (1762, *ad loc.*):

Qui enim in malis et infortuniis abditus detinetur, ut vector in navi, quomodo is intelligat quales affectus ex forte felici oriri solent, fortunam ipse scilicet faventem nunquam, quantum quidem ex dictis constat, expertus?

Come si può avere ἐπιστήμη della buona e della cattiva sorte, se ci si trova immersi nella sola condizione negativa? Come si può, dal solo lato della privazione, avere nozione della presenza? La domanda, che sembra voler stringere nella morsa sillogistica la verità poetica, ha avuto invece fortuna, a giudicare dalle molteplici risposte. In particolare:

- 1) preferire, come Heath stesso, ἔμπειρος a ἔμπορος, intendendo l'esperienza dei mali come condizione non preclusiva dell'esperienza dei beni;
- 2a) sostituire, come SCHÜTZ (1823), κακῶν con βίου:

Quomodo enim ii, qui κακῶν tantum ἔμποροι aut ἔμπειροι sunt, deinde dicantur intelligere, eos qui secundis rebus utantur, eandem sortem sibi sperare? [...] Vitam vero hominum saepe cum navigatione comparari, notius est, quam ut moneri debeat. Quodsi βίου legamus, sententia pulcherrime finitur, absque ulla translationis inconsequentia [...]

2b) modificare, come WEIL (1884), βροτοῖσιν in βροτεῖον e scambiare la posizione di questa parola del verso seguente con il nostro κακῶν. Sull'intervento di Weil si tornerà in seguito.

Contro la prima soluzione, credo sia sufficiente obiettare che la condizione dell'ἔμπορος è, per definizione, transitoria: non vi è quindi alcuna necessità logica, come pretende Heath, che chi è κακῶν ἔμπορος di conseguenza non abbia mai avuto esperienza del bene. Anzi, semmai, il κακῶν ἔμπορος, a seguire l'etimologia, è colui che si trova di passaggio, colui che transita sulla nave dei mali.

Contro la seconda soluzione, nelle sue due varianti, si può seguire il ragionamento espresso nelle *Adnotationes* di Hermann, che tuttavia preferisce ἔμπειρος:

Est autem κακῶν multo fortius quam βίου [e si può aggiungere: et quam βροτεῖον]. Etenim si queras quid hoc sit, βίου, nihil reperias aliud quam

8. Commento a tre luoghi scelti

utramque fortunam. Atqui quum prius de adversa loquatur poeta, *κακῶν* ponit, quod et accommodatum ad hanc sententiam et per se ipsum etiam melius est. Secundam enim fortunam omnes plus minus notam habent, adversa vero non omnes affligit. Itaque qui hanc sunt experti, hi demum recte de vita iudicare possunt, quippe sine dubio etiam meliore aliquando fortuna usi: quod non praedicaveris de felicibus, fuisse eos aliquando etiam miseros. Quare non opus fuit ut utriusque fortunae gnaros commemoraret.

La riserva di natura semantica e sintattica riguarda la reggenza del genitivo.

Dice BLOMFIELD (1823, *ad loc.*):

Ceteri, cum Scholiasta, *ἔμπορος*, quae quidem vox, quando in hoc sensu ponitur, cum genitivo non construitur, namque in Eurip. *Hippol.* 964. *ἔμπορον βίου* est mercatricem vitae. Sed non est dubium quin verum sit *ἔμπειρος*, [...]

GROENEBOOM (1930) è ancora più categorico²⁶:

bei der Lesart der besten Handschriften *κακῶν* - *ἔμπορος* würde jeder Grieche an einen malorum mercator denken und niemals auf den Gedanken kommen, daß Aischylos hier die Vorstellung navigans calamitatibus habe ausdrücken wollen; ohne Zweifel ist also *ἔμπειρος* vorzuziehen, vgl. z.B. Soph. *Ant.* 1191 *κακῶν γὰρ οὐκ ἄπειρος οὖσ' ἀκούσομαι*, O.K. 751 *οὐ γάμων ἔμπειρος* [...]

Forse fuori contesto qualunque Greco avrebbe potuto pensare a *malorum mercator*²⁷, come dice Groeneboom, ma seguito a breve termine dalla metafora *κλύδων κακῶν* del verso successivo, poteva ben essere inteso *navigans*, come del resto fa lo scolio che lo rende con *ἐμπλέων*.

L'argomento a sfavore resterebbe quindi l'attestazione in tragedia del solo significato di *mercator*, quando *ἔμπορος* è costruito col genitivo. Tuttavia, vale la pena vagliare gli argomenti a favore del significato di *navigans*, *vector* e simili. Innanzitutto, le prime attestazioni di *ἔμπορος* propendono per il significato di «one who goes on ship board as a passenger» (*LSJ*, *s.v.*; si veda anche GARVIE (2009, *ad loc.*)), «celui qui voyage sur un bateau qui ne lui appartient pas, passenger» (Chantraine): è il caso di β 319 *οὐ γὰρ νηὸς ἐπήβολος οὐδ' ἐρετῶν* e ω 299-301

²⁶Riportiamo il passo nella traduzione tedesca.

²⁷Sulla possibile *abusio* del termine si veda SCHUURSMA (1932).

ποῦ δαὶ νηὺς ἔστηκε θεή, ἣ σ' ἤγαγε δεῦρο
 ἀντιθέους θ' ἑτάρους· ἦ ἔμπορος εἰλήλουθας
 νηὸς ἐπ' ἀλλοτρίης, οἱ δ' ἐκβήσαντες ἔβησαν·

È doveroso comunque far presente che Heath ricava da passi come *ω* 299-301 l'idea che sia necessario dire *ἔμπορος ἐπὶ νηὸς* e non sia consentito invece dire *ἔμπορος νηός*. Inoltre, le altre due attestazioni eschilee del termine, purtroppo costruite entrambe senza genitivo, sono impiegate comunque nell'accezione di *viator*: P. Ox. 2162, I, I: 20 *ἔμπορον* (-ων Fraenk.) *κωλύτορα - ὥρα δ' ἔμπορους καθιέναι ἄγκυραν*. Gli aggettivi composti con *-πόρος* presenti in Eschilo sono *ἄπορος* (PV 904), *ἀντίπορος*, (*Supp.* 544, *Pers.* 67) *εὐπορος* (*Supp.* 470), più il derivato *πόριμος* (PV 904). Notevoli sono i giochi etimologici nei quali si trovano coinvolti: PV 904 *Ἀπόλεμος ὅδε γ' ὁ πόλεμος, ἄπορα πόριμος Supp.* 544 *ἀντίπορον ... πόρον*.

Non è da dimenticare poi che il termine *πόρος*, *nomen actionis* da *πείρω*, che occorre in Eschilo 32 volte (ed. Murray) e ben 9 volte solo nei *Persiani*, si può costruire col genitivo del luogo da attraversare.

CHANTRAINE (1999) dà come etimologia di *ἔμπορος*, fra l'altro supportata anche dall'*Etim. M.*, *ἐν πόρω ὄν*, grazie alla quale non è difficile pensare, sfruttando il gioco etimologico, una dipendenza di *κακῶν* da *ἔμπορος*, intendendo tutta la frase: “chiunque si trovi nella condizione di attraversamento dei mali” o, appunto, “chiunque sia un *viator malorum*”. Per finire, si può prendere in considerazione il sostantivo composto *συνέμπορος*, che occorre col significato di *comes*, in senso proprio in *Cho.* 208, *Supp.* 939 (*συνέμποροι σέθεν*), *Cho.* 713, o figurato in *Cho.* 733 (*σοι ξυνέμποροι*). Roussel, forse in modo un po' forzato, a ulteriore sostegno della possibilità che *ἔμπορος*, nel significato di «passager» e non di «mercator» possa reggere il genitivo del luogo, e non quindi della cosa che si commercia, riporta il passo di Ar. *Ra.* 395: *θεὸν τὸν συνέμπορον τῆσδε τῆς χορείας* e traduce: «le dieu qui prend part à cette danse». Dato che *συν-* porta il composto a reggere il dativo di compagnia, è evidente che la reggenza del genitivo insiste tutta sulla seconda parte del composto, cioè *-ἔμπορον*.

8. Commento a tre luoghi scelti

Sembra di un qualche riguardo anche l'osservazione di Teuffel, che accoglie nel suo testo *ἐμπερής*:

eine auch bei Sophokles (fr. 412 D.) vorkommende Form für *ἔμπορος*.
Äschylos wählte diese Form, weil er im 5. Fusse dem Spondeus den Jambus vorzog.

In effetti, tenendo conto dell'osservazione iniziale di Roussel sulla improbabilità del mutamento da *ἔμπειρος* ad *ἔμπορος*, il difficile *ἐμπερής* potrebbe stare all'origine di entrambi i rami della tradizione, corrotti per ragioni diverse. Tuttavia, anche questa osservazione della preferenza per il giambo al quinto piede può essere preso come ulteriore prova della bontà di *ἔμπορος* rispetto a *ἔμπειρος*.

Passiamo ora al secondo intervento testuale, cioè *αἰὲν ἄνεμον* in luogo di *αἰεὶ δαίμων*'. È da rimarcare il fatto che sia Murray, sia Page, sia West accolgano questa congettura, mentre gli editori non anglosassoni adottano maggiore prudenza. Anche in questo caso, a dire il vero, in assenza di esigenze metriche, non sembra ci sia il carattere di necessità. In effetti WEIL (1884), quando propone questa congettura, la giustifica molto rapidamente:

La répétition du mot *δαίμων* ne doit sans doute pas être attribuée au poète. Je crois que *αἰεὶ δαίμων*' provient de *αἰὲν ἄνεμον*. Ainsi s'explique aussi le génitif *τύχης* qui ne peut guère être gouverné par *οὐριεῖν*.

Bisogna tener conto tuttavia che Weil aveva proposto interventi molto pesanti a questo passo, come si può vedere già dalla sua precedente edizione dei Persiani, WEIL (1867):

[...] *ἐπίσταται κακῶν μὲν ὡς ὄτω κλύδων
καινῶν ἐπέλθη, πάντα δειμαίνειν φιλεῖ
ὅταν δ' ὁ δαίμων εὐροῇ* [...]

Così giustifica le sue scelte:

Tres offensiones habet vulgata: quod rerum humanarum, non solum adversae fortunae peritus dicendus erat, quod *κακῶν* moleste repetitur, quod *δειμαίνειν* non timendi, ut solet, sed terrendi vim habet. Itaque verba leviter immutata sic transposui ut res humanae in adversas secundasque dividerentur, pro altero *κακῶν* scripsi *καινῶν*, ὅταν mutavi in ὄτω.

Evidentemente la congettura αἰὲν ἄνεμον si trova sulla stessa linea. Semanticamente affini ad ἄνεμον sono le congetture πνεύματ' ... τύχης di Wecklein e αἰὲν ἀτμόν ... τύχης di Scheer. Il mutamento di τύχης della paradosi in accusativo, singolare o plurale, ha trovato il favore di molti editori, per due ragioni: offrire a οὐριεῖν un complemento oggetto ed evitare la ridondanza di δαίμον' ... τύχης. Roussel spiega bene quale potrebbe essere stata la genesi del mutamento da τύχας a τύχης: «Le copiste a pris ce mot pour un génitif dorien intempestif et a écrit τύχης». Tuttavia lo scolio spiega οὐριεῖν con l'intransitivo οὐριοδρομεῖν e il composto κατουρίζω, di significato affine quando è usato transitivamente, è usato intransitivamente in Soph. Tr. 827 col significato traslato di “andare a buon fine”. Per quanto riguarda δαίμον' ... τύχης DE ROMILLY (1974) fa notare:

le tour δαίμονα τύχης est assez inhabituel, mais on trouve dans Pindare τύχη δαίμονος (Ol., VIII, 67).

Questa si potrebbe quindi intendere come ipallage di un'espressione comune attestata. Se δαίμον' non fosse autentico, si potrebbe spiegare l'errore come generato dall'inserimento nel testo di una glossa interlineare a τὸν αὐτόν.

Per concludere, credo sia opportuno prendere in considerazione le figure presenti ai vv. 598-606 di M:

figurae per adiectionem

ripetizione: κακῶν ... κακῶν, ὅταν ... ὅταν, πάντα ... πάντα, δαίμων ... δαίμον' (potenziato dalla consonanza con δειμαίνειν)

richiamo etimologico: φίλοι ... φιλεῖ, φόβου ... ἐκφοβεῖ

allitterazione: κλύδων / κακῶν in enjambement, αὐτόν αἰὲν

Da notare inoltre l'insistenza della metafora marinara²⁸: κακῶν ... ἔμπορος, κλύδων (notevole anche la consonanza di εὐροῆ ... οὐριεῖν. In Q², come nota marginale, è addirittura scritto εὐροεῖν per οὐριεῖ.

figurae per detractionem e figurae per immutationem

ellissi del soggetto e costruzione del plurale col singolare: βροτοῖσιν (da cui ricavare un soggetto plurale) ... φιλεῖ.

ἀπὸ κοινοῦ: δειμαίνειν φιλεῖ ... πεποιθέναι

²⁸Si veda VAN NES (1963).

8. Commento a tre luoghi scelti

figurae per ordinem

chiasmo (v. 604): ἐν ὄμμασιν ... φαίνεται ... βοᾷ δ' ἐν ὣσιν

Il quadro, credo, dovrebbe mostrare la prevalenza di *figurae per adiectionem* ma soprattutto come queste figure siano solidali fra di loro a creare, in questi versi, una sorta di rondò che si apre con la menzione dei mali in generale e si chiude con la menzione dei mali del momento. Per tutte queste ragioni, sembra prudente conservare non solo il testo della *paradosi*, ma di **M** in particolare.

8.3 *Pers.* 704

Il v. 704 offre un caso paradigmatico dei sottili equilibri che si stabiliscono fra valutazione dell'*usus* di un autore su base lessicale e non semantica e applicazione del criterio della *lectio difficilior*. Dario, dopo aver cercato inutilmente di ottenere informazioni concise dal coro, si rivolge ad Atossa con queste parole (vv. 703-708):

ἀλλ' ἐπεὶ δέος παλαιὸν σοὶ φρενῶν ἀντίσταται,
τῶν ἐμῶν λέκτρων γεραιὰ ξύννομ', εὐγενὲς γύναι,
κλαυμάτων λήξασα τῶνδε καὶ γόων σαφές τί μοι
λέξον. ἀνθρώπεια δ' ἄν τοι πῆματ' ἂν τύχοι βροτοῖς.
πολλὰ μὲν γὰρ ἐκ θαλάσσης, πολλὰ δ' ἐκ χέρσου κακὰ
γίγνεται θνητοῖς, ὁ μᾶσσων βίωτος ἦν ταθῆ πρόσω.

Così nella edizione di Murray, che segue la tradizione principale. Tuttavia Page solleva un dubbio, e pur accogliendo nel testo *γύναι* scrive in apparato:

γύναι: δάμαρ I, fort. recte

e l'ipotesi acquista ancora più forza per West, che accoglie nel testo *δάμαρ* e porta a sostegno anche un luogo parallelo:

δάμαρ I (cf. fr. 383): *γύναι* cett.

Il testo del fr. 383 R. dice infatti:

Ἥρα τελεία, Ζηνὸς εὐναία δάμαρ

Del resto, il vocabolo occorre anche in *PV* 560 δάμαρτα, 767 δάμαρτος, 834 δάμαρ, ma l'incertezza sulla paternità del *Prometeo*²⁹ spinge probabilmente West a tacere questi passi.

È da notare che nel frammento la costruzione sintattica di δάμαρ con il nome dello sposo al genitivo è conforme all'uso omerico: le cinque occorrenze del termine nei poemi epici sono infatti le seguenti: *Ε* 503 ἡ Προμάχοιο δάμαρ, *δ* 126 Πολύβοιο δάμαρ, *Γ* 122 Ἀντηγορίδαο δάμαρτι *ν* 290, *ω* 125 Ὀδυσσῆος δὴν οἰχομένοιο δάμαρτα. Nel *Prometeo* il v. 834 ἡ Διὸς κλεινῆ δάμαρ ci offre lo stesso tipo di costruzione, al contrario del v. 767 Ἡ πρὸς δάμαρτος ἐξανίσταται θρόνων· e del v. 560 ἄγαγες Ἡσιόναν πιθὼν δάμαρτα κοινόλεκτρον, dove κοινόλεκτρον può ricordare il nostro τῶν ἐμῶν λέκτρων ... ξύννομ', per quanto la circostanza descritta sia piuttosto diversa. Ma questi motivi sono troppo deboli per determinare la scelta della variante, tanto più che, se ci si estende agli altri autori di teatro, troviamo un parallelo sia alla *iunctura* con il termine più elevato: εὐγενῆς δάμαρ di Euripide (*Hipp.* 26), che tuttavia usa il termine con una frequenza tale da non permettere confronti con gli altri due tragici³⁰, sia alla *iunctura* con il termine meno elevato: εὐγενῆς γυνή di Soph. *El.* 257, il quale tuttavia non è semanticamente pertinente con il nostro passo, dato che la diversa condizione di Elettra elimina la polisemia di γυνή, valida invece per Atossa. In Ar. *Thesm.* 330 troviamo εὐγενεῖς γυναῖκες, che in questo caso è invece carico di tutta la *vis polemica* delle donne riunite in assemblea. È bene quindi cercare altri elementi a favore delle rispettive varianti.

Il codice **I** (Athous Ἰβήρων 209) è un codice del XIII/XIV secolo investigato per la prima volta da Dawe, sulla scorta del quale West esprime il suo giudizio dicendo:

Est autem proximus post M bonitate. Sunt quae solus servavit (Pers. 704. Prom. 502. 716. cf. ad Sept. 565. 804. 915. 923), sunt quae solus praeter M (Sept. 86. 308. cf. ad Prom. 903), quocum etiam errores quosdam

²⁹Altri editori, fra cui Murray stesso, accolgono nel testo dell'*Agamennone* al v. 1495 e al corrispondente v. 1519 dell'antistrofe l'integrazione congetturale di Enger δαμεις <δάμαρτος>, ovviamente non utilizzabile come luogo parallelo.

³⁰Una ricerca rapida sul TLG dà oltre un centinaio di occorrenze in Euripide contro le 4 di Eschilo e le 9 di Sofocle.

8. Commento a tre luoghi scelti

communes habet; scholiaque vetera interdum plenius quam M reddit, cum in ceteris desint.

Tuttavia dei tre presunti luoghi che dovrebbero preservare le lezioni genuine, uno è il nostro sotto esame e gli altri due sono interventi minimi, rispettivamente $\tau\epsilon$ in luogo di $\delta\acute{\epsilon}$, sanabile per congettura, e $\pi\rho\acute{o}\sigma\pi\lambda\alpha\tau\omicron\iota$ in luogo di $\pi\rho\acute{o}\sigma\pi\lambda\alpha\sigma\tau\omicron\iota$ degli altri codici, che potrebbe anche essere un “felice errore” di copiatura. Dal medesimo ipoarchetipo di **I** provengono altri due codici del XV secolo, uno dei quali, **Ba**, è privo del testo dei *Persiani*, mentre l’altro è siglato con Δ (Mosqu. gr. 508). Su questi codici il giudizio di West è severo:

Ba Δ ipsos raro laudo, quoniam una eorum utilitas est ut e concordia IBa Δ vel IBa vel I Δ lectiones hyparchetypi a reciperemus.

E del resto Δ nel caso in esame non è di aiuto, visto che reca $\gamma\upsilon\nu\eta$ come tutti gli altri codici e non ci fornisce quindi elementi meccanici per stabilire se $\delta\acute{\alpha}\mu\alpha\rho$ si trovasse nell’ipoarchetipo. Un’ipotesi plausibile è che comunque $\delta\acute{\alpha}\mu\alpha\rho$ non si trovasse nel corpo del testo dell’ipoarchetipo, dato che in quel caso Δ in tutta probabilità lo avrebbe copiato. Non è escluso invece che si trovasse forse a margine o in interlinea come *varia lectio* nell’ipoarchetipo.

Sulla qualità di $\delta\acute{\alpha}\mu\alpha\rho$ il giudizio di Dawe è perentorio:

*$\delta\acute{\alpha}\mu\alpha\rho$ is a totally unexpected variant found only in I. It is a word which implies respect, and as such is very much in place as the last word in a line the whole of which constitutes a dignified address to the Queen. Nor is it difficult to see that the highly poetic word would quite naturally receive the obvious gloss $\gamma\acute{\iota}\nu\alpha\iota$. Could $\delta\acute{\alpha}\mu\alpha\rho$ have been a deliberate and arbitrary ‘improvement’? All that we can reply is that, if it is, it is a very extreme example. There is no doubt at all that the *obvious* interpretation is that $\delta\acute{\alpha}\mu\alpha\rho$ has been uniquely preserved by I. Any other view, at this stage (i.e. before the formulation of a theory of manuscript descent), cannot escape being called wilfully perverse.*

Non sorprende infatti che, tanto negli *scholia vetera* quanto nei *recentiora*, nei tre luoghi citati del *Prometeo* sia proprio $\gamma\upsilon\nu\eta$ a glossare $\delta\acute{\alpha}\mu\alpha\rho$, così come $\gamma\upsilon\nu\eta$ viene usato nella spiegazione etimologica da parte di Triclinio del composto $\delta\upsilon\sigma\delta\acute{\alpha}\mu\alpha\rho\tau\omicron\varsigma$ impiegato nell’*Agamennone*³¹. Si potrebbe pensare quindi di trovarsi

³¹Il termine è usato da Cassandra ai vv. 1318-1319:

di fronte all'ordinario meccanismo di trivializzazione del testo, che porterebbe ad accogliere *δάμαρ* su suggerimento di Dawe, Page e West.

Ma l'idea che un codice rechi una sola lezione genuina e sia l'unico a preservarla induce a cercare spiegazioni alternative. In effetti abbiamo visto che **I** ha accesso a un insieme di scoli più ricco di **M** stesso e gli scoli recano informazioni di varia natura, ad esempio luoghi paralleli o interpretazioni, e non soltanto semplificazioni lessicali, che possono finire nel testo. Un buon candidato sarebbe proprio l'euripideo *εὐγενῆς δάμαρ*. Ma c'è un sottile motivo per cui il comune e polisemico *γυνή* potrebbe essere stato glossato con l'«highly poetic word» *δάμαρ*: la volontà di sciogliere l'ambiguità in modo inequivocabile, usando cioè un termine elevato ma dal significato univoco di “sposa legittima”. Questa esigenza potrebbe essere nata dalla constatazione che Dario parla di Atossa oppure si rivolge a lei usando espressioni che lasciano sempre aperta l'ambiguità: *ἄκοιτιν* al v. 684, *τῶν ἐμῶν λέκτρων ... ξύννομ'* e il discusso *γυνή* al v. 704.

La plausibilità di questa ipotesi trova conferma in uno scolio marginale al codice **Ya** (Vindob. phil. gr. 197) del 1413, che precisa *ἄκοιτιν* del v. 684 scrivendo:

ἤγουν τὴν ἐμὴν δάμαρτα

Bene osserva BELLONI (1994, *ad loc.*):

Dario si rivolge ad Atossa con vocativi di onore, fra i quali *γύναι* (cfr. i vv. 156, 623). [...] *γύναι* appartiene alla titolatura di Atossa nei due passi citati.

Alla luce di queste considerazioni, ci si trova d'accordo con l'osservazione di ROUSSEL (1960):

*ὅταν γυνή γυναικὸς ἀντ' ἐμοῦ θάνη,
ἀνὴρ τε δυσδάμαρτος ἀντ' ἀνδρὸς πέσῃ.*

Lo scolio dice: *ὁ κακὴν ἔχων γυναικα.*

8. Commento a tre luoghi scelti

Ici, c'est plutôt *noble femme* que *noble épouse*, qui ferait double emploi avec *ξύννομε* [...] Le roi, comme on voit, est roi et aussi époux. La reine est reine et mère, mais à peu près pas épouse.

9

Conclusioni

Giunti al termine di questo libro, sembra opportuno ricordare brevemente i punti salienti di ciascun capitolo e i risultati principali.

L'introduzione si è concentrata sulle definizioni di filologia collaborativa e cooperativa, osservando come le tecnologie digitali e i metodi computazionali non solo facilitino la collaborazione fra studiosi che partecipano a progetti comuni, ma permettano anche la cooperazione fra comunità che costruiscono risorse e strumenti condivisi in modo aperto, al fine di essere usati ed estesi per scopi diversi da quelli per cui sono stati creati inizialmente. Si è proposta inoltre la distinzione fra filologia digitale, volta principalmente alla creazione e alla marcatura di risorse, e filologia computazionale, volta invece principalmente allo sviluppo di strumenti di analisi.

L'acquisizione del testo digitale, tema del secondo capitolo, è sempre più efficace grazie al miglioramento delle tecnologie di riconoscimento ottico dei caratteri supportate da modelli linguistici che permettono di incrementare l'accuratezza del testo digitalizzato. La fase di digitalizzazione è il vero collo di bottiglia che ha rallentato lo sviluppo della filologia digitale e, di conseguenza, della filologia computazionale. Si sta lasciando per così dire, l'età dell'innocenza in cui varianti, congetture e letteratura secondaria erano ignorate, e si sta entrando nell'età della maturità, che vede crescere i primi studi di filologia "nativi digitali". Si passa quindi dall'interesse per il solo testo dell'opera antica, come nei primi

9. Conclusioni

corpora di testi greci e latini, all'attenzione per l'edizione critica nella sua interezza, comprensiva di testo, apparato e altri elementi paratestuali quali prefazione, introduzione, indici, etc.

L'analisi di apparati e repertori, discussa nel terzo capitolo, ha fornito l'occasione per esplorare le modalità di marcatura manuale suggerite dalla *Text Encoding Initiative* per gli apparati critici ed estendibile anche agli apparati di congetture. Si è visto come la marcatura applicata a campioni basati su repertori di congetture basati su diverse edizioni di riferimento sia utile per ragionare su come è strutturata l'informazione in essi contenuta.

Il trattamento automatico di apparati e repertori, argomento del quarto capitolo, permette di estrarre informazioni strutturate da documenti semi-strutturati, in modo da ridurre i tempi (e di conseguenza i costi) rispetto alla marcatura manuale. Particolare attenzione si è posta all'allineamento delle informazioni di apparato rispetto alle edizioni di riferimento, al fine di creare indici e concordanze che tengano conto della variantistica e al fine di comparare aspetti linguistici e stilistici delle unità testuali allineate.

I modelli e le risorse per l'analisi sintattica, presi in considerazione nel quinto capitolo, hanno fornito l'occasione per ragionare sull'impiego di *treebanks* nel campo degli studi classici. Si è visto come diversi modelli di rappresentazione sintattica si siano succeduti nel tempo e come attualmente si converga verso la *Dependency Grammar*.

I modelli e le risorse lessico-semantici, trattati nel sesto capitolo, hanno mostrato come sia possibile individuare, rappresentare e trattare le relazioni fra parole, fra parole e concetti e fra concetti e concetti. Il paradigma strutturalista e il paradigma cognitivista sono stati discussi e messi a confronto.

L'esplorazione degli spazi semantici, illustrata nel settimo capitolo, si è focalizzata su tre aspetti peculiari dei *corpora* di testi letterari antichi: la dimensione relativa piccola, la diacronicità e la ripartizione in generi letterari. La pertinenza delle relazioni semantiche individuate, infatti, è fortemente condizionata da questi parametri.

Il commento a tre luoghi scelti dei *Persiani* di Eschilo, che ha occupato l'ottavo capitolo, si è posto l'obiettivo di riportare l'attenzione del filologo digitale e computazionale sui metodi tradizionali della filologia classica e in particolare sul metodo del commento lineare, fatto attraverso il confronto delle fonti primarie e della letteratura secondaria.

Si cercherà ora di illustrare come tale modo di operare della filologia classica possa essere integrato e potenziato dalla filologia digitale e computazionale.

L'incontro fra studi filologici e informatica può essere inteso in modo debole oppure in modo forte. Nel primo caso, lo scopo è di mettere a disposizione del filologo copie digitali delle fonti primarie e secondarie necessarie allo studio dei testi classici, al fine di ridurre la necessità di recarsi fisicamente presso le biblioteche. Per ottenere questo obiettivo più debole, sono sufficienti immagini di buona qualità, metadati precisi e una discreta accuratezza dell'OCR.

Nel secondo caso invece la pagina di testo non deve essere soltanto leggibile dagli studiosi, ma deve essere anche *machine actionabile*, per fare in modo che il testo di diverse edizioni possa essere cercabile e fruibile per analisi linguistiche e stilistiche. Per ottenere questo obiettivo più forte, è necessaria un'accuratezza elevata dell'OCR e modelli robusti per rappresentare il testo non solo come sequenza lineare di caratteri, ma come struttura complessa che includa la variantistica. I filologi digitali devono quindi mettere a disposizione dei filologi computazionali ampie collezioni di documenti acquisiti in modo accurato per poter essere analizzati linguisticamente e stilisticamente, senza rinnegare la tradizione di studi che, nel corso dei secoli, ha affinato i propri metodi per la costituzione del testo e l'interpretazione del suo contenuto. La filologia collaborativa e cooperativa si pone in continuità, non in rottura con questa tradizione.

Abitualmente i filologi valutano le congetture consultando luoghi paralleli dello stesso autore, dello stesso genere letterario o dello stesso periodo, usando indici e concordanze. Le applicazioni di *text retrieval*, come ad esempio *Diogenes*¹, hanno molto facilitato la ricerca di questi parallelismi, ma non hanno modificato la metodologia tradizionale, basata sull'individuazione sistematica di somiglianze

¹<https://community.dur.ac.uk/p.j.heslin/Software/Diogenes>

9. Conclusioni

lessicali, non necessariamente semantiche. In effetti, i parallelismi lessicali possono essere determinati da una corrispondenza esatta fra forme flesse o forme lemmatizzate. Le similarità semantiche, al contrario, non sempre trovano consenso unanime.

Come si è visto al capitolo 7, i metodi esplorativi permettono di mostrare termini fortemente associati in base ai contesti d'uso all'interno di ampie collezioni testuali, suggerendo allo studioso delle ipotesi d'indagine da verificare poi in profondità. Nell'era digitale, il filologo ha quindi l'opportunità di passare dal *close reading* cui è abituato al *distant reading*² dei nuovi metodi d'indagine su grandi quantità di dati da osservare con uno sguardo d'insieme, per tornare poi al *close reading* con maggiore consapevolezza o qualche intuizione aggiuntiva.

Se la classificazione semantica delle varianti è importante per studiare come le glosse interlineari possano essere entrate nel testo o come lo scriba abbia banalizzato la lezione originale, d'altra parte la classificazione delle congetture può aiutare il filologo a studiare il modo di operare dei filologi che lo hanno preceduto, le loro tendenze a correggere per similarità o per contrasto. Verrà illustrato un esempio per chiarire.

Si torni al passo già visto al capitolo 8, cioè *Pers.* 133-139 (edizione di West):

λέκτρα δ' ἀνδρῶν πόθῳ
πίμπλαται δακρύμασιν
Περσίδες δ' ἀβροπενθείς ἐκάσ-
τα πόθῳ φιλάνορι
τὸν αἰχμᾶεντα θούρον εὐνατῆρ' ἀποπεμφαμένα
λείπεται μονόζυξ.

Come si è già detto, la ripetizione di *πόθῳ*, alla distanza di pochi versi diede origine a molte congetture, come si vede nel repertorio di Wecklein:

λέκτρα δ' ἀμῖν μάταν Enger. πόνῳ Pauw, σπάνει olim, postea ἔρω Heimsoeth, ὀδῶ Oberdick.

²Si veda MORETTI (2015). I *Voyant Tools* sono una buona dimostrazione di come si possa passare agevolmente da una modalità all'altra (<http://voyant-tools.org>).

λέκτρα	δ'	ἀνδρῶν	πόθῳ	
		ἀμῖν	μάταν	Enger
			πόνῳ	Pauw
			σπάνει	Heimsoeth
			ἔρω	Heimsoeth
			ὀδῶ	Oberdick

Tabella 9.1: Allineamento di congetture

Dopo l'allineamento automatico delle congetture con il testo dell'edizione di riferimento, come mostrato in Tab. 9.1, ciascuna parola allineata a *πόθῳ* viene confrontata con la lista delle prime cinque associazioni semantiche relative a *πόθος*, generata con lo strumento *associate* di cui si è parlato al capitolo 7. Tale lista contiene *ποθέω* (desiderare), *φίλτρον* (filtro d'amore), *ἔρως* (amore), *νυμφίδιος* (nuziale) e *ἔρος* (desiderio). In questo modo il fatto che Heimsoeth, dopo un primo tentativo di emendamento con *σπάνει*, formulò una congettura costituita da un sinonimo del termine che intende eliminare, è messo in evidenza in modo automatico e quindi può essere trovato non solo quando si cercano informazioni su questo passo specifico (*close reading*), ma anche se l'obiettivo della ricerca ha un grado di astrazione maggiore, come l'individuazione di tutte le congetture contenenti sinonimi (*close*), con l'accortezza di verificare poi, caso per caso, se lo studioso che ha formulato la congettura ha dichiarato esplicitamente il motivo della sua scelta (di nuovo *close reading*, questa volta sulla letteratura secondaria). Basta confrontare³ HEIMSOETH (1862, p. 61) con HEIMSOETH (1865, p. 136) per avere conferma che lo studioso ha formulato la prima congettura in base all'interpretazione di uno scolio e la seconda congettura in base alla similarità semantica, confermata da Esichio, fra *πόθος* ed *ἔρως*.

³Entrambe le opere citate di Heimsoeth sono disponibili su *Internet Archive*.

Bibliografia

- A. ABEILLÉ. *Les nouvelles syntaxes: Grammaires d'unification et analyse du Français*. A. Colin, Paris 1993.
- A. ABEILLÉ (ed.). *Trebanks. Building and Using Parsed Corpora*. Kluwer Academic Publishers, Dordrecht - Boston - London 2003.
- F. ADRADOS. *Nueva Sintaxis del Griego antiguo*. Gredos, Madrid 1992.
- F. ADRADOS. *La semántica en el diccionario Griego-Español*. In *Cien años de investigación semántica*, 99–110. Ed. Clásicas, Madrid 2000.
- F. ADRADOS, E. GANGUTIA, J. L. FACAL, C. S. AYBAR (ed.). *Introducción a la lexicografía griega*. Inst. Antonio de Nebrija, Madrid 1977.
- J. ALLEN. *Natural Language Understanding*. Addison Wesley, New York 1995, 2 ed.
- T. ALLISON. *Aeschylean Stylistics: A Study of Linguistic Variation*. Tesi di dottorato, University of Michigan 2003.
- J. ALLSOPP. *Microformats: Empowering Your Markup for Web 2.0*. New York, NY 2007.
- B. ALMAS, M. BEAULIEU. *Developing a New Integrated Editing Platform for Source Documents in Classics. Literary and Linguistic Computing* 28.4 (2013), 493–503.

BIBLIOGRAFIA

- D. APOLLON, C. BELISLE. *Digital Critical Editions*. University of Illinois Press, Chicago 2014.
- H. BAAYEN. *Analyzing Linguistic Data - A Practical Introduction to Statistics Using R*. Cambridge, MA 2008.
- A. BABEU. *Rome Wasn't Digitized in a Day: Building a Cyberinfrastructure for Digital Classicists*. Council on Library and Information Resources, Washington, DC 2011.
- D. BAMMAN, F. MAMBRINI, G. CRANE. *An Ownership Model of Annotation: The Ancient Greek Dependency Treebank*. In *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories*. Milan 2009.
- L. BELLONI. *Eschilo. I Persiani*. Vita e pensiero, Milano 1994, 2 ed.
- É. BENVENISTE. *Le vocabulaire des institutions indo-européennes - 1. économie, parenté, société*. Les Éditions de Minuit, Paris 1969.
- M. BERTI, Y. BIZZONI, F. BOSCHETTI, G. CRANE, R. DEL GRATTA, T. YOUSEF. *Ancient Greek WordNet Meets the Dynamic Lexicon: the Example of the Fragments of the Greek Historian*. In *Proceedings of the Eighth Global WordNet Conference GWC-2016*, 34–38. Global WordNet Association, Bucharest 2016.
- A. BIES, M. FERGUSON, K. KATZ, R. MACINTYRE. *Bracketing Guidelines for Treebank II Style - Penn Treebank Project*. Rap. Tecn., Linguistic Data Consortium. University of Pennsylvania 1995.
- Y. BIZZONI, F. BOSCHETTI, H. DIAKOFF, R. DEL GRATTA, M. MONACHINI, G. CRANE. *The Making of Ancient Greek WordNet*. In N. CALZOLARI, K. CHOUKRI, T. DECLERCK, H. LOFTSSON, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK, S. PIPERIDIS (ed.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. European Language Resources Association (ELRA), Reykjavik 2014.

- Y. BIZZONI, R. D. GRATTA, F. BOSCHETTI, M. REBOUL. *Enhancing the Accuracy of Ancient Greek WordNet by Multilingual Distributional Semantics*. In *Proceedings of CLiC-it 2015*. 2015.
- C. BLOMFIELD (ed.). *Persae, not. et gloss. adj.* C.J. Blomfield. Wöller, Leipzig 1823, 3 ed.
- B. BOGURAEV, J. PUSTEJOVSKY (ed.). *Corpus Processing for Lexical Acquisition*. Cambridge, MA 1996.
- A. BÖHMOVÁ, J. HAJIČ, E. HAJIČOVÁ, B. HLADKÁ. *The Prague Dependency Treebank*. In ABEILLÉ (2003), cap. 7, 103–127.
- F. BOSCHETTI. *Studio degli spazi semantici con strumenti informatici come metodo esplorativo per la valutazione di congetture*. *Bollettino dei Classici* 3.s. 30 (2009), 41–53.
- F. BOSCHETTI. *Acquisizione e Creazione di Risorse Plurilingui per gli Studi di Filologia Classica in Ambienti Collaborativi*. In M. AGOSTI, F. TOMASI (ed.), *Collaborative Research Practices and Shared Infrastructures for Humanities Computing*, 55–67. CLEUP, Padova 2014.
- F. BOSCHETTI, M. ROMANELLO, A. BABEU, D. BAMMAN, G. CRANE. *Improving OCR Accuracy for Classical Critical Editions*. In M. AGOSTI, J. BORBINHA, S. KAPIDAKIS, C. PAPTAEODOROU, G. TSAKONAS (ed.), *Research and Advanced Technology for Digital Libraries – 13th European Conference, ECDL 2009, Corfu, Greece, Sept./Oct. 2009, Proceedings*, 156–167. 2009.
- F. BOSCHETTI, R. DEL GRATTA, M. LAMÉ. *Computer Assisted Annotation of Themes and Motifs in Ancient Greek Epigrams: First Steps*. In R. BASILI, A. LENCI, B. MAGNINI (ed.), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 Vol. I*, 83–86. Pisa 2014.
- F. BOTHE (ed.). *Aeschlyi Tragoediae*, ed. Fr. H. Bothe. Volumen prius: *Prometheus vincitus, Supplices, Septem adversus Thebas, Persae*. Hahn, Leipzig 1831.

BIBLIOGRAFIA

- A. BOZZI. *Character Recognition and the Linguistic Spelling Checker: an integrated technique*. In A. BOZZI (ed.), *Computer-aided recovery and analysis of damaged text documents*, 161–186. Bologna 2000.
- A. BOZZI. *Verso una filologia computazionale: la prima Euroconferenza della European Science Foundation*. *Euphrosyne* n.s. 32 (2004a), 127–138.
- A. BOZZI. *Postfazione a Zampolli Antonio, Filologia e informatica: le origini della filologia computazionale*. *Euphrosyne* n.s. 32 (2004b), 21–24.
- A. BOZZI, A. NIKOLOVA, G. CAPPELLI, G. GIULIANI. *Il trattamento delle varianti nello spoglio elettronico di un testo. Una prova sui Carmina di Claudiano. Materiali e Discussioni per l'Analisi dei Testi Classici* 16 (1986), 155–179.
- T. BRANTS, W. SKUT, H. USZKOREIT. *Syntactic annotation of a german newspaper corpus*. In ABEILLÉ (2003), cap. 5, 73–87.
- C. BREMOND (ed.). *Recherches rhétoriques*. Éditions du Seuil, Paris 1994.
- J. BRESNAN. *The Mental Representation of Grammatical Relations*. The MIT Press, Cambridge, MA 1982.
- J. BRESNAN. *Lexical-Functional Syntax*. Blackwell, Oxford 2001.
- H. BROADHEAD. *Aeschylus. The Persae*, edited with an introduction, critical notes and commentary by H.D. Broadhead. Cambridge University Press, Cambridge 1960.
- N. BRÖKER. *A Projection Architecture for Dependency Grammar and How it Compares to LFG*. In M. BUTT, T. H. KING (ed.), *Proceedings of the LFG98 Conference*. 1998.
- D. BUZZETTI. *Rappresentazione digitale e modello del testo*. In C. ET AL. (ed.), *Il ruolo del modello nella scienza e nel sapere*, 127–161. Accademia Nazionale dei Lincei, Roma 1999.

- M. BÜCHLER, P. BURNS, M. MÜLLER, E. FRANZINI, G. FRANZINI. *Towards a Historical Text Re-use Detection*. In *Text Mining*, 221–238. Springer International Publishing 2014.
- I. CALVINO. *Una pietra sopra*, cap. Cibernetica e fantasmi (Appunti sulla narrativa come processo combinatorio), 199–219. O. Mondadori, Milano 1995.
- A. CANDIO. *Pregare e maledire: Aesch. Ch. 145 s.*, vol. 32. 2014.
- H. CECOTTI, A. BELAÏD. *Hybrid OCR combination approach complemented by a specialized ICR applied on ancient documents*. In *8th International Conference on Document Analysis and Recognition*, 1045–1049. 2005.
- G. CELANO. *Guidelines for the Annotation of the Ancient Greek Dependency Treebank 2.0*. https://github.com/PerseusDL/treebank_data/edit/master/AGDT2/guidelines 2014.
- P. CHANTRAINE. *Dictionnaire étymologique de la langue grecque*. Klincksieck, Paris 1999.
- N. CHOMSKY. *Lectures on Government and Binding - The Pisa Lectures*. Foris, Dordrecht 1981.
- N. CHOMSKY. *The Minimalist Program*. The MIT Press, Cambridge, MA 1995.
- F. CIOTTI. *La rappresentazione digitale del testo: il paradigma del markup e i suoi sviluppi*. In L. PERILLI, D. FIORMONTE (ed.), *La macchina nel tempo: studi di informatica umanistica in onore di Tito Orlandi*. Le lettere, Firenze 2011.
- V. CITTI. *Il linguaggio religioso e liturgico nelle tragedie di Eschilo*. Zanichelli, Bologna 1962.
- V. CITTI. *Eschilo e la lexis tragica*. Hakkert, Amsterdam 1994.
- A. CIULA, F. STELLA (ed.). *Digital Philology and Medieval Texts*. Ospaletto–Pisa 2007.

BIBLIOGRAFIA

- E. COSERIU. *Teoría del lenguaje y lingüística general - Cinco estudios*. Gredos, Madrid 1973, 3 ed.
- G. CRANE. *Generating and Parsing Classical Greek. Literary and Linguistic Computing* 6.6 (1991), 243–245.
- G. CRANE, D. BAMMAN, L. CERRATO, A. JONES, D. MIMNO, A. PACKEL, D. SCULLEY, G. WEAVER. *Beyond Digital Incunabula: Modeling the Next Generation of Digital Libraries*. In *10th European Conference on Research and Advanced Technology for Digital Libraries, volume 4172 of Lecture Notes in Computer Science*, 353–366. Springer, Berlin 2006.
- G. CRANE, B. SEALES, M. TERRAS. *Cyberinfrastructure for Classical Philology. Digital Humanities Quarterly* 3.1 (2009), 1–27.
- M. CROCHEMORE, C. HANCART, T. LECROQ. *Algorithms on Strings*. Cambridge University Press 2007.
- C. DALITZ, R. BASTON. *Optical Character Recognition with the Gamera Framework*. In C. DALITZ (ed.), “*Document Image Analysis with the Gamera Framework*”. *Schriftenreihe des Fachbereichs Elektrotechnik und Informatik. Hochschule Niederrhein* 8, 53–65. Shaker Verlag, Maastricht 2009.
- M. DALRYMPLE (ed.). *Semantics and Syntax in Lexical Functional Grammar*. The MIT Press, Cambridge, MA 1999.
- R. DAWE. *The collation and investigation of the manuscripts of Aeschylus*. Cambridge University Press, Cambridge, U.K. 1963.
- R. DAWE. *Repertory of conjectures on Aeschylus*. Brill, Leiden 1965.
- J. DE ROMILLY. *La crainte et l'angoisse dans le théâtre d'Eschyle*. Les Belles Lettres, Paris 1958.
- J. DE ROMILLY (ed.). *Eschyle. Les Perses - Édition, introduction et commentaire par un groupe de normaliens sous la direction de Jacqueline de Romilly*. Presses universitaires de France, Paris 1974.

- E. DEGANI. *Il mostro di Irvine*. *Eikasmos* 3 (1992), 277–278.
- R. DEL GRATTA, F. BOSCHETTI, A. DEL GROSSO, F. KHAN, M. MONACHINI. *Cooperative Philology on the Way to Web Services: The Case of the CoPhiWordNet Platform*. In D. L. Y. MURAKAMI (ed.), *Worldwide Language Service Infrastructure*, 173–187. Springer, Berlin 2016.
- R. DELMONTE. *Linguistic and inferential processes in text analysis by computer*. Unipress, Padova 1992.
- M. DRISCOLL, E. PIERAZZO (ed.). *Digital Scholarly Editing: Theories and Practices*, vol. 4 di *Digital Humanities Series*. Open Book Publishers 2016.
- O. DUCROT, J. SCHAEFFER (ed.). *Nouveau dictionnaire encyclopédique des sciences du langage*. Éditions du Seuil, Paris 1995.
- H. EBELING. *Lexicon Homericum*. Teubner, Lipsia 1885.
- U. ECO. *Sugli specchi*. Bompiani, Milano 1985.
- U. EGLI. *Die Anwendbarkeit der generativen Grammatik auf das Altertumswissenschaft*. In G. REDARD (ed.), *Indogermanische und allgemeine Sprachwissenschaft*. Reichert, Wiesbaden 1973.
- S. EVERT, M. BARONI. *zipfR: Word frequency distribution in R*. In *Proceedings of the Demo and Poster Sessions of ACL 2007*, 29–32. East Stroudsburg, PA 2007.
- É. ÉVRARD. *Une informatisation de la syntaxe de dépendance en Latin*. In M. L. D. LONGRÉE (ed.), *Actes du Ve Colloque de Linguistique latine*, Cahiers de l'Institut de Linguistique de Louvain, 115–126. Louvain-la-Neuve 1989.
- P. FABBRI, G. MARRONE (ed.). *Semiotica in nuce - Vol. II. Teoria del discorso*. Meltemi, Roma 2000.
- C. FABRE, A. LENCI. *Distributional Semantics Today. Introduction to the special issue*. *TAL* 56.2 (2015), 7–20. URL <https://hal.archives-ouvertes.fr/hal-01259695>.

BIBLIOGRAFIA

- C. FELLBAUM (ed.). *Wordnet - An Electronic Lexical Database*. Cambridge, MA 1998a.
- C. FELLBAUM (ed.). *Wordnet - An Electronic Lexical Database*. The MIT Press, Cambridge, MA 1998b.
- S. FENG, R. MANMATHAN. *A Hierarchical, HMM-based Automatic Evaluation of OCR Accuracy for a Digital Library of Books*. In *Proceedings of the Joint Conference in Digital Libraries*, 109–118. 2006.
- C. FILLMORE. *The need for a frame semantics in linguistics*. *Statistical Methods in Linguistics* 12 (1997), 5–29.
- J. FODOR. *Three reasons for not deriving 'kill' from 'cause to die'*. *Linguistic inquiry* 1 (1970), 429–438.
- W. FRANCIS, H. KUČERA. *Manual of information to accompany a standard corpus of present-day edited American English for use with digital computers*. Rap. Tecn., Brown University, Providence, RI 1964.
- G. FRANZINI, M. BERTI, E. FRANZINI, G. CRANE, G. CELANO. *L'Open Philology Project dell'Università di Lipsia: Per una filologia "sostenibile" in un mondo globale*. In M. AGOSTI, F. TOMASI (ed.), *Collaborative Research Practices and Shared Infrastructures for Humanities Computing*, 151–162. CLEUP, Padova 2013.
- D. FROGER. *La critique des textes et son automatisation*. Dunod, Paris 1968.
- A. GARVIE (ed.). *Aeschylus: Persae*. Oxford University Press, Oxford 2009.
- B. GATOS, G. LOULLOUDIS, N. STAMATOPOULOS. *Greek Polytonic OCR Based on Efficient Character Class Number Reduction*. In *Proceedings of the 2011 International Conference on Document Analysis and Recognition (ICDAR 11)*. IEEE, Piscataway 2011. URL <http://dx.doi.org/10.1109/ICDAR.2011.233>.

- G. GAZDAR, C. MELLISH. *Natural Language Processing in Prolog - An Introduction to Computational Linguistics*. Addison Wesley, Wokingham 1989.
- M. GIORDANO. *Góos aretós tra maledizione e vendetta di sangue: un saggio di analisi semantica*. *AION(filol)* 20 (1998), 59–77.
- A. GREIMAS. *Sémantique structurale*. Presses Universitaires de France, Paris 1995, 2 ed.
- M. GRIFFITH. *The vocabulary of the Prometheus Bound*. *CQ* 34 (1984), 282–291.
- P. GROENEBOOM. *Aeschylus. Persae*. Wolters, Groningen 1930.
- D. A. GROSSMAN, O. FRIEDER. *Information Retrieval: Algorithms and Heuristics*. Springer, New York 2004.
- C. GULICK (ed.). *The Deipnosophists of Athenaeus*. Loeb Classical Library, London – New York, NY 1951, 2 ed.
- M. GUPTA, N. JACOBSON, E. GARCIA. *OCR Binarization and Image Pre-processing for Searching Historical Documents*. *Pattern Recognition* 40.2 (2007), 389–397. URL <http://dx.doi.org/10.1016/j.patcog.2006.04.043>.
- B. HEATH. *Notae sive lectiones ad tragicorum Graecorum veterum Aeschyli, Sophoclis, Euripidis quae supersunt dramata deperditorumque reliquias*. e typogr. Clarendoniano, Oxonii 1762.
- F. HEIMSOETH. *Die indirecte Ueberlieferung des aeschylischen Textes*. Max Cohen & Sohn, Bonn 1862.
- F. HEIMSOETH. *Kritische Studien zu den Griechischen Tragikern. Erste Abtheilung. Eine nothwendige Ergänzung der kritischen Methode*. Max Cohen & Sohn, Bonn 1865.
- G. HERMANN (ed.). *Aeschyli tragoediae*. Weidmann, Lipsiae 1852.

BIBLIOGRAFIA

- M. HILPERT, S. GRIES. *Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition*. *Literary and Linguistic Computing* 24.4 (2009), 385–401.
- L. HJELMSLEV. *Omkring Sprogteoriens Grundlaeggelse*. The linguistic circle of Copenhagen, Copenhagen 1993.
- E. HOBBS, I. LAWRENCE. *An Outline of a Transformational-Generative Grammar of Hellenistic Greek*. Berkeley 1978, 2 ed.
- G. ITALIE. *Index Aeschyleus*. E.J. Brill, Leiden 1964.
- P. JUDET DE LA COMBE. *L'Agamemnon d'Eschyle - Le texte et ses interprétations*. Pr. Univ. de Lille, Lille 1982.
- M. JUILLARD, X. LUONG. *De nouveaux arbres pour un nouveau corpus*. *RISSH* 30.
- J. JUSTESON, S. KATZ. *Co-occurrences of Antonymous Adjectives and Their Contexts*. *Computational Linguistics* 17.1 (1991), 1–19.
- G. KAIBEL (ed.). *Deipnosophistarum Libri XV*. Teubner, Lipsiae 1887.
- M. KAY. *Parsing in Functional-Unification Grammar*. In D. DOWTY, L. KARTTUNEN, A. ZWICKY (ed.), *Natural Language Understanding and Logic Programming*. Cambridge Univ. Press, Cambridge, MA 1985.
- KLOWSKI. *Was ist die generative Transformationsgrammatik und welche Bedeutung könnte sie für den altsprachlichen Unterricht haben?* *AU* 14 (1971), 5–19.
- G. KONDRAK. *Algorithms for Language Reconstruction*. Tesi di dottorato, University of Toronto, Toronto 2002. URL <http://www.cs.ualberta.ca/~kondrak/papers/thesis.pdf>.
- A. KONTOSTATHIS, W. POTTENGER. *A Framework for Understanding Latent Semantic Indexing (LSI) Performance* – Paper online. <http://webpages.ursinus.edu/akontostathis/KontostathisIP&M.pdf> 2003.

- D. KORZENIEWSKI. *Studien zu den Persern des Aischylos*. *Helikon* 7 (1967), 27–62.
- O. KOZAR. *Towards Better Group Work: Seeing the Difference between Cooperation and Collaboration*. *English Teaching Forum* 2 (2010), 16–23.
- R. KÜHNER, B. GERTH. *Ausführliche Grammatik der griechischen Sprache*. Verlag Hahnsche Buchhandlung, Hannover-Leipzig 1898, 3 ed.
- H. KURZOVÁ. *Transformational Grammar and Greek Syntax*. *Eirene* 10 (1972), 21–52.
- G. LAKOFF. *Metaphors we live by*. Chicago Univ. Press, Chicago - London 1980.
- G. LAKOFF. *Women, Fire, and Dangerous Things - What Categories Reveal about the Mind*. Univ. of Chicago Press, Chicago - London 1987.
- H. LAUSBERG. *Handbuch der Literarischen Rhetorik*. F. Steiner Verlag, Stuttgart 1990, 3 ed.
- A. LENCI. *Distributional semantics in linguistic and cognitive research. A foreword*. *Rivista di Linguistica* 1 (2008), 1–30.
- M. LOMBARDO. *Habrosýne e habrá nel mondo greco arcaico*. In *Modes de contacts et processus de transformation dans les sociétés antiques - Actes du colloque de Cortone (24-30 mai 1981)*, 1077–1103. 1983.
- G. LONGOBARDI. *Lezioni di sintassi generale e comparata*. Editoria Universitaria, Venezia 1991.
- W. LUND, E. RINGGER. *Improving Optical Character Recognition through Efficient Multiple System Alignment*. In *Joint Conference for Digital Libraries*. 2009.
- W. LUND, D. KENNARD, E. RINGGER. *Why Multiple Document Image Binarizations Improve OCR*. In *Proceedings of the 2nd International Workshop*

BIBLIOGRAFIA

- on Historical Document Imaging and Processing (HIP '13)*, 86–93. ACM, New York 2013. URL <http://dx.doi.org/10.1145/2501115.2501126>.
- J. LYONS. *Semantics*. Cambridge Univ. Press, Cambridge - New York - Melbourne 1977.
- M. MAGRIS, M. MUSACCHIO, L. REGA, F. SCARPA (ed.). *Manuale di terminologia - Aspetti teorici, metodologici e applicativi*. Hoepli, Milano 2002.
- F. MAMBRINI. *L'Ancient Greek Dependency Treebank. Un nuovo strumento per lo studio della lingua greca*. *Lexis* 51–70.
- F. MAMBRINI. *Thucydides 1.89-118: A Multi-layer Treebank*. *CHS Research Bulletin* 1.2. URL http://nrs.harvard.edu/urn-3:hinc.essay:MambriniF.Thucydides_1.89-118_Multi-layer_Treebank. 2013.
- F. MAMBRINI, M. PASSAROTTI. *Non-Projectivity in the Ancient Greek Dependency Treebank*. In *Proceedings of the Second International Conference on Dependency Linguistics, DepLing 2013, August 27-30, 2013, Prague, Czech Republic*, 177–186. 2013.
- F. MAMBRINI, M. PASSAROTTI. *Subject-Verb Agreement with Coordinated Subjects in Ancient Greek*. *Journal of Greek Linguistics* 16 (2016), 87–116.
- M. MARCUS, M. MARCINKIEWICZ, B. SANTORINI. *Building a large annotated corpus of English: the Penn Treebank*. *Comput. Linguist.* 19.2 (1993), 313–330.
- A. MARTINET. *Syntaxe générale*. A. Colin, Paris 1985.
- M. MARTÍNEZ HERNÁNDEZ. *Semántica del Griego antiguo*. Ediciones Clásicas, Madrid 1997.
- M. MARTÍNEZ HERNÁNDEZ (ed.). *Cien años de investigación semántica: de Michel Bréal a la actualidad - Actas del congreso internacional de semántica, Universidad de La Laguna, 27-31 de octubre de 1997*. Ed. Clásicas, Madrid 2000.

- A. MARZAL, E. VIDAL. *Computation of normalized edit distance and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence* 15.9 (1993), 926–932.
- L. MASSA POSITANO. *Sopra alcuni versi dei “Persiani” di Eschilo. Dioniso* n.s. 12 (1949), 63–67.
- P. MASTANDREA. *Gli archivi elettronici di ‘Musisque deoque’. Ricerca intertestuale e cernita fra varianti antiche. In ZURLI E MASTANDREA (2009), 41–72.*
- G. MATINO. *La sintassi di Eschilo. M. D’Auria Editore, Napoli 1998.*
- P. MAZON. *Sur deux passages d’Eschyle et sur une formule d’Homère. REG* 63 (1950), 11–19.
- A. MEINEKE (ed.). *Athenaei Deipnosophistae. Teubner, Lipsiae 1858.*
- E. MINCHIN. *Homer and the resources of memory - Some applications of cognitive theory to the Iliad and the Odyssey. Oxford University Press, Oxford - New York 2001.*
- L. MONDIN. *Appunti per una critica (inter)testuale della poesia latina. In ZURLI E MASTANDREA (2009), 73–106.*
- R. MORDENTI. *Informatica e critica dei testi. Bulzoni editore 2001.*
- F. MORETTI. *Distant Reading. Verso, London 2015.*
- H. MORIER. *Dictionnaire de poétique et de rhétorique. Presses Universitaires de France, Paris 1981.*
- G. MURRAY. *Aeschyli septem quae supersunt tragoediae. Clarendon Press, Oxford 1955.*
- A. NAMBOODIRI, P. NARAYANAN, C. JAWAHAR. *On Using Classical Poetry Structure for Indian Language Post-Processing. In 9th International Conference*

BIBLIOGRAFIA

- on Document Analysis and Recognition*, vol. 2, 1238–1242. IEEE Computer Society 2007.
- G. NAVARRO, M. RAFFINOT. *Flexible Pattern Matching in Strings*. Cambridge University Press 2002.
- J. NIVRE, M. DE MARNEFFE, F. GINTER, Y. GOLDBERG. *Universal Dependencies v1: A Multilingual Treebank Collection*. In *Proceedings of LREC*. 2016.
- P. NORVIG. *Paradigms of Artificial Intelligence Programming - Case Studies in Common Lisp*. Morgan Kaufmann Publishers, San Francisco, CA 1992.
- M. B. O'DONNELL. *Corpus Linguistics and the Greek of the New Testament*. Sheffield Phoenix, Sheffield, U.K. 2005.
- T. PACKER. *Performing Information Extraction to Improve OCR Error Detection in Semi-structured Historical Documents*. In *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, 67–74. ACM, New York 2011. URL <http://doi.acm.org/10.1145/2037342.2037354>.
- D. PAGE (ed.). *Aeschylus. Septem quae supersunt Tragoedias* edidit Denys Page. Clarendon Press, Oxford 1972.
- F. PALEY (ed.). *Aeschylus*. Translated into English prose by F.A. Paley. Second edition revised. Deighton, Bell and C., Cambridge 1871.
- T. PARR. *The Definitive ANTLR 4 Reference*. The Pragmatic Programmers, Raleigh, NC 2014.
- M. PASSAROTTI. *One Hundred Years Ago. In Memory of Father Roberto Busa SJ*. In F. MAMBRINI, M. PASSAROTTI, C. SPORLEDER (ed.), *Proceedings of The Third Workshop on Annotation of Corpora for Research in the Humanities*, 15–24. The Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia 2013.

- C. PAVESE. *Semantematica della poesia corale greca*. *Belfagor* 23 (1968), 389–430.
- C. PAVESE. *I temi e i motivi della lirica corale ellenica* - Introduzione, analisi e indice semantematici - Alcmane, Simonide, Pindaro, Bacchilide. Istituti Editoriali e Poligrafici Internazionali, Pisa 1997.
- J. PENNINGTON, R. SOCHER, C. MANNING. *GloVe: Global Vectors for Word Representation*. In *Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. 2014. URL <http://www.aclweb.org/anthology/D14-1162>.
- H. PERDICOYIANNI-PALÉOLOGUE. *Bibliographie retrospective: sémantique grecque (1981-2000)*. *Lexis* 21 (2003), 1–26.
- J. PÉREZ MARTEL. *Semántica y hermenéutica en los escolios y glosas de las tragedias de Esquilo*. In M. M. HERNÁNDEZ (ed.), *Cien años de investigación semántica*, 1161–1174. Ed. Clásicas, Madrid 2000.
- R. PFISTER. *Thesen zu Linguistik und Sprachunterricht. Anregung. Zeitschrift für die Höhere Schule* 15 (1969), 374–376.
- R. PFISTER. *Schlusswort zu den Thesen zu Linguistik und Sprachunterricht. Gymnasium* 79 (1972), 314–330.
- P. PICKERING. *Repetitions and their removal by the copyists of greek Tragedy. GRBS* 41.2 (2000), 123–139.
- E. PIERAZZO. *A rationale of digital documentary editions. Literary and Linguistic Computing* 26.4 (2011), 463–477.
- E. PIERAZZO. *Digital Scholarly Editing : Theories, Models and Methods*. Digital Research in the Arts and Humanities. Ashgate, Farnham Surrey 2015.
- M. PIOTROWSKI. *Natural Language Processing for Historical Texts*, vol. 5.2 di *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers 2012.

BIBLIOGRAFIA

- A. RADFORD. *Transformational Grammar - A first Course*. Cambridge Univ. Press, Cambridge, MA 1988.
- F. RASTIER. *Sémantique interprétative*. Presses Universitaires de France, Paris 1996, 2 ed.
- F. RASTIER. *Sémantique et recherches cognitives*. Presses Universitaires de France, Paris 2001, 2 ed.
- F. RASTIER, M. CAVAZZA, A. ABEILLÉ. *Sémantique pour l'analyse - De la linguistique à l'informatique*. Masson, Paris 1994.
- M. REYNAERT. *Non-interactive OCR Post-correction for Giga-Scale Digitization Projects*. In *CICLing 2008*, LNCS 4919, 617–630. 2008a.
- M. REYNAERT. *All, and Only, the Errors: More Complete and Consistent Spelling and OCR-Error Correction Evaluation*. In *Proceedings of the 6th International Conference on Language Resources and Evaluation 2008*, 1867–1872. Marrakech, Morocco 2008b.
- L. REYNOLDS, N. WILSON. *Scribes and Scholars – A Guide to the Transmission of Greek and Latin Literature*. Oxford, U.K. 1991, 3 ed.
- L. RICHARDSON. *Persae 135. Hermathena* 89 (1957), 73–75.
- G. RIGO. *Eschyle. Opera et fragmenta omnia – Index verborum – Listes de fréquence*. Liège 1999.
- C. RINGLSTETTER, K. SCHULZ, S. MIHOV, K. LOUKA. *The same is not the same - postcorrection of alphabet confusion errors in mixed-alphabet OCR recognition*. In *8th International Conference on Document Analysis and Recognition*, vol. 1, 406–410. 2005.
- B. ROBERTSON, F. BOSCHETTI. *Large-Scale Optical Character Recognition of Ancient Greek*. *Mouseion* 3.s. 14 (2017), 341–360.

- B. ROBERTSON, C. DALITZ, F. SCHMITT. *Automated Page Layout Simplification of Patrologia Graeca*. In *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage (DATECH '14)*, 167–172. ACM, New York 2014. URL <http://dx.doi.org/10.1145/2595188.2595213>.
- M. RODDA, M. SENALDI, A. LENCI. *Panta rei: Tracking Semantic Change with Distributional Semantics in Ancient Greek*. *Italian Journal of Computational Linguistics* 11–24.
- H. RODRÍGUEZ, S. CLIMENT, P. VOSSEN, L. BLOKSMA, W. PETERS, A. ALONGE, F. BERTAGNA, A. ROVENTINI. *The Top-Down Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology*. *Computers and the Humanities* 32 (1998), 117–152.
- G. RONCAGLIA. *La quarta rivoluzione: sei lezioni sul futuro del libro*. Laterza, Roma 2010.
- L. ROUSSEL (ed.). *Eschyle. Les Perses - Texte, traduction, commentaire*. Presses Universitaires, Montpellier 1960.
- H. RUBENSTEIN, J. GOODENOUGH. *Contextual correlates of synonymy*. *Comm. ACM* 8.10.
- P. SAHLE. *Digitale Editionsformen: Textbegriffe und Recodierung (Teil 3). Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, vol. 9 di *Schriften des Instituts für Dokumentologie und Editorik*. 2013.
- M. SAHLGREN. *The Word-Space Model*. Tesi di dottorato, University of Stockholm, Stockholm 2006.
- M. SAHLGREN. *The distributional hypothesis*. *Rivista di Linguistica* 20.1 (2008), 33–54.
- G. SALTON, A. WONG, C. YANG. *A vector space model for automatic indexing*. *Communications of ACM* 18.11 (1975), 613–620.

BIBLIOGRAFIA

- R. SCHANK, R. ABELSON. *Scripts Plans Goals and Understanding - An Inquiry into Human Knowledge Structures*. L. Erlbaum Ass., Hillsdale, NJ 1977.
- D. SCHMIDT. *Nominalizing Transformations in Hellenistic Greek*. Graduate Theol. Union, Berkeley 1979.
- D. SCHMIDT, R. COLOMB. *A Data Structure for Representing Multi-version Texts Online*. *International Journal of Human-Computer Studies* 67(6.4) (2009), 497–514. URL <http://dx.doi.org/10.1016/j.ijhcs.2009.02.001>.
- C. SCHÜTZ (ed.). *Aeschyli Dramata*. G. and W.B. Whittaker, London 1823.
- J. SCHUURSMA. *De poetica vocabulorum abusione apud Aeschylum*. Paris, Amsterdam 1932.
- E. SCHWYZER. *Griechische Grammatik*. Beck, München 1990, 6 ed.
- G. SEMERANO. *Le origini della cultura europea - 2, Dizionari etimologici: basi semitiche delle lingue indeuropee - Dizionario della lingua greca*. Leo S. Olschki Editore, Firenze 1994.
- A. SIDERAS. *Aeschylus Homericus. Untersuchungen zu den Homerismen der aeschyleischen Sprache*. Vandenhoeck & Ruprecht, Göttingen 1971.
- W. SKUT, B. KRENN, T. BRANTS, H. USZKOREIT. *An Annotation Scheme for Free Word Order Languages*. In *Proceedings of the Fifth Conference on Applied Language Processing*, 27–28. Washington 1997.
- B. SNELL. *Die Entdeckung des Geistes Studien zur Entstehung des europäischen Denkens bei den Griechen*. Classen, Hamburg 1955, 3 ed.
- M. SPENCER, C. HOWE. *Collating texts using progressive multiple alignment*. *Computer and the Humanities* 37.1 (2003), 97–109.
- C. SPERBERG-MCQUEEN, L. BURNARD (ed.). *Guidelines for Text Encoding and Interchange*. Humanities Computing Unit - Univ. of Oxford, Oxford 2002.

- U. SPRINGMANN, A. LÜDELING. *OCR of historical printings with an application to building diachronic corpora: A case study using the RIDGES herbal corpus*. *CoRR* abs/1608.02153. URL <http://arxiv.org/abs/1608.02153>.
- L. B. K. A. SQUITIER. *Thesaurus Linguae Graecae Canon of Greek Authors and Works*. Oxford 1990, 3 ed.
- A. STANZIONE, G. RE, G. MUGELLI, A. TADDEI, F. BOSCHETTI, R. DEL GRATTA. *Homeric Greek WordNet: costruire una risorsa lessico-semantica fra ricerca e didattica*. In F. BOSCHETTI (ed.), *AIUCD 2016: Book of Abstracts*, 129–132. AIUCD, Firenze 2017.
- L. STERLING, E. SHAPIRO. *The art of Prolog*. MIT Press, Cambridge, MA 1994, 2 ed.
- G. STEWART, G. CRANE, A. BABEU. *A New Generation of Textual Corpora*. In *Joint Conference of Digital Libraries 2007*, 356–365. 2007.
- M. TERRAS, J. NYHAN, E. VANHOUTTE (ed.). *Defining Digital Humanities: a Reader*. London – New York 2016.
- L. TESNIÈRE. *Éléments de syntaxe structurale*. Klincksieck, Paris 1988, 5 ed.
- J. TRIER. *Aufsätze und Vorträge zur Wortfeldtheorie*. Mouton, The Hague - Paris 1973.
- R. TRUJILLO. *Algunas observaciones acerca del referente*. In M. M. HERNÁNDEZ (ed.), *Cien años de investigación semántica*, 121–141. Ediciones Clásicas, Madrid 2000.
- J. VAN BEUSEKOM, F. SHAFAIT, T. BREUL. *Automated OCR Ground Truth Generation*. In *9th International Conference on Document Analysis and Recognition*. 2007.
- D. VAN NES. *Die maritime Bildersprache des Aischylos*. J.B. Wolters, Groningen 1963.

BIBLIOGRAFIA

- E. VANHOUTTE. *Defining Electronic Editions: A Historical and Functional Perspective*. In W. MCCARTY (ed.), *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*, Digital Humanities, 119–144. Open Book Publishers 2010.
- M. E. WALL, A. RECHTSTEINER, L. ROCHA. *Singular Value Decomposition and Principal Component Analysis*. Norwell, MA 2003.
- N. WECKLEIN (ed.). *Aeschyli fabulae cum lectionibus et scholiis codicis Medicei et in Agamemnonem codicis Florentini ab Hieronymo Vitelli denuo collatis*, edidit N. Wecklein. S. Calvary, Berlin 1885. 2 voll.
- N. WECKLEIN (ed.). *Aeschyli fabulae cum lectionibus et scholiis codicis Medicei et in Agamemnonem codicis Florentini ab Hieronymo Vitelli denuo collatis*, edidit N. Wecklein - Vol. II: Partis II auctarium: *Appendix propagata*. S. Calvary, Berlin 1893.
- H. WEIL (ed.). *Aeschyli quae supersunt tragoediae*. Recensuit, adnotationem criticam et exegeticam adjecit Henricus Weil in Facultate Litterarum Vesontina professor. Ricker, Gissae 1867. 2 voll.
- H. WEIL (ed.). *Aeschyli tragoediae*. Edidit Henricus Weil. Teubner, Leipzig 1884.
- M. WEST. *Studies in Aeschylus*. Teubner, Stuttgart 1990.
- M. WEST (ed.). *Aeschylus. Tragoediae cum incerti poetae Prometheus*. Teubner, Lipsiae 1998.
- N. WHITE. *Training Tesseract for Ancient Greek OCR*. *Eutypon* 28–29 (2012), 1–11.
- T. WINOGRAD. *Understanding Natural Language*. Academic Press, New York 1972.
- I. YALNIZ, R. MANMATHA. *A Fast Alignment Scheme for Automatic OCR Evaluation of Books*. In *Proceedings of the 2011 International Conference on*

- Document Analysis and Recognition (ICDAR)*, 754–758. IEEE, Piscataway, NJ 2011. URL <http://dx.doi.org/10.1109/ICDAR.2011.157>.
- T. YOUSEF, C. PALLADINO. *iAligner: A tool for syntax-based intra-language text alignment*. In F. BOSCHETTI (ed.), *AIUCD2016: Book of Abstracts*, 201–206. AIUCD, Firenze 2017.
- L. ZHUANG, X. ZHU. *An OCR Post-processing Approach Based on Multi-knowledge*. In *9th International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, 346–352. 2005.
- H. ZINSMEISTER, J. KUHN, B. SCHRADER, S. DIPPER. *TIGER Transfer - From LFG Structures to the TIGER Treebank*. Rap. Tecn., IMS, University of Stuttgart, Stuttgart 2001.
- G. ZIPF. *Human behavior and the principle of least effort*. Cambridge 1949.
- L. ZURLI, P. MASTANDREA (ed.). *Poesia Latina, Nuova E-Filologia – Opportunità per l'editore e per l'interprete*. Roma 2009.

