

# FARE LINGUISTICA APPLICATA CON LE *DIGITAL HUMANITIES*

a cura di

Jacopo Saturno - Lorenzo Spreafico

studi AltLA **14**

AltLA

**studi AltLA 14**

FARE LINGUISTICA APPLICATA  
CON LE *DIGITAL HUMANITIES*

a cura di  
JACOPO SATURNO - LORENZO SPREAFICO

Milano 2022

L'AItLA pubblica una collana di monografie e di collettanee sui diversi temi della linguistica applicata. I manoscritti vengono valutati con i consueti processi di revisione di pari per assicurarne la conformità ai migliori standard qualitativi del settore. I volumi sono pubblicati nel sito dell'associazione con accesso libero a tutti gli interessati.

*Comitato scientifico*

Cecilia Andorno, Giuliano Bernini, Simone Ciccolone, Silvia Dal Negro, Maria Elena Favilla, Anna De Meo, Laura Gavioli, Antonietta Marra, Natacha S.A. Niemants, Elena Nuzzo, Jacopo Saturno, Lorenzo Spreafico, Marilisa Vitale.

© 2022 AItLA - Associazione Italiana di Linguistica Applicata  
Via Cartoleria, 5  
40100 Bologna - Italy  
email: [info@aitla.it](mailto:info@aitla.it)  
sito: [www.aitla.it](http://www.aitla.it)



Edizione realizzata da  
Officinaventuno  
Via Flli Bazzaro, 18  
20128 Milano - Italy  
email: [info@officinaventuno.com](mailto:info@officinaventuno.com)  
sito: [www.officinaventuno.com](http://www.officinaventuno.com)

ISSN: 2724-4830

ISBN edizione cartacea: 978-88-97657-50-7

ISBN edizione digitale: 978-88-97657-51-4

edizione digitale distribuita secondo la licenza Creative Commons CC BY NC ND



# Indice

Introduzione	5
FRANCESCO CUTUGNO - EMANUELA CAMPISI Linguistica applicata alle esperienze di cultura digitale. Analisi multimodale nel progetto CHROME	9
CHRISTOPH DRAXLER Automatic Transcription of Spoken Language Using Publicly Available Web Services	27
VALERIA CARUSO - ROBERTA PRESTA Approcci digitali alla lessicografia: sviluppo e valutazione di un dizionario per smartphone	49
SIMONE CICCOLONE - GIULIA GROSSO Macchine parlanti o macchine apprendenti? Osservazioni preliminari sull'uso dell'etero-ripetizione in produzioni di chatbot e nell'interlingua	69
VALENTINA DE IACOVO - MARCO PALENA La variazione prosodica in italiano: l'utilizzo di un chatbot Telegram per la didattica assistita per apprendenti di italiano L2 e nella valutazione linguistica delle conoscenze disciplinari	85
GIOVINA ANGELA DEL ROSSO - SILVIA BRAMBILLA L'accuratezza della trascrizione ASR sul parlato non-standard. L'italiano nell'OH Portal	99
MARTA MAFFIA - MASSIMO PETTORINO - ANNA DE MEO - ROSA DE MICCO - GIOACCHINO TEDESCHI - ALESSANDRO TESSITORE Ritmo e malattia di Parkinson. Prime riflessioni sull'annotazione automatica del parlato di pazienti in stadi iniziali	117
LUISA REVELLI - ANTONIO MASTROPAOLO - DANIELE PAOLO RADICIONI La <i>sottodeterminazione</i> nei testi giuridici: verso un'analisi linguistico- computazionale	131
FLAVIA SCIOLETTE - EMILIANO GIOVANNETTI Un modello per domarli tutti: verso una rappresentazione del testo come esplicitazione di documento, lingua e contenuto	145
LORENZO SPREAFICO Fare linguistica applicata con le <i>digital humanities</i> : interfaccia di metodo o di comodo uso?	159
Indice autori	173



## La *sottodeterminazione* nei testi giuridici: verso un'analisi linguistico-computazionale<sup>3</sup>

### *Abstract*

This contribution introduces a theoretical-methodological paradigm framed in an interdisciplinary research project at the intersection among law, linguistics and natural language processing, focused on the legislative and regulatory production of Valle D'Aosta Region, as a limited though autonomous body of rules.

Main target of the project is the notion of underdetermination, that is the property of juridical utterances to allow language to convey different, alternative interpretations. The experimental methods employed for the field survey are illustrated, along with the applications of digital humanities used to devise a descriptive model of phenomenology and linguistic realities of underdetermination.

The final aim of the project is to describe and to track underdetermination in the many different forms it takes in the wide range of the different text typologies encompassed in the juridical field.

### 1. *Inquadramento*

*Nell'applicare la legge non si può ad essa attribuire altro senso che quello fatto palese dal significato proprio delle parole secondo la connessione di esse, e dalla intenzione del legislatore.*

“Disposizioni sulla legge in generale”, Art. 12, comma I.

Per quanto in parte mitigato dal riferimento alla *ratio iuris* (l'*intenzione del legislatore*) e da successive determinazioni che prefigurano in alternativa all'interpretazione letterale una serie di possibilità ermeneutiche sempre ad essa gerarchicamente subordinate, la previsione del *Codice civile* posta in epigrafe si affida al convincimento di matrice cognitivista che la lingua possa rappresentare in modo univoco e inequivocabile il valore significativo (*sensu*) ricavabile dallo specifico contenuto semantico (*significato proprio*) delle singole unità linguistiche presenti (*parole*) e dalla loro concatenazione nel contesto morfosintattico (*secondo la connessione di esse*).

---

<sup>1</sup> Università della Valle d'Aosta.

<sup>2</sup> Università degli Studi di Torino.

<sup>3</sup> I contenuti qui esposti sono frutto delle riflessioni teoriche e delle scelte metodologiche condivise in piena collaborazione dagli autori. Per quanto riguarda la stesura del testo, sono stati elaborati in comune i paragrafi 1, 2 e 7; Antonio Mastropaolo è autore del paragrafo 3; Daniele Paolo Radicioni del paragrafo 5; Luisa Revelli dei paragrafi 4 e 6 e della revisione generale del testo.

Il contributo che qui presentiamo si concentra sulle circostanze in cui il contenuto espresso da un enunciato normativo non presenta invece le caratteristiche di evidenza attese, e pone anzi problemi di decodifica tali da impedire un'interpretazione univoca. Si propone, più in particolare, di delineare un quadro teorico che consenta di identificare le condizioni che determinano la possibilità di un enunciato giuridico di prestarsi a diverse e alternative interpretazioni plausibili.

Mutuando il termine dall'ambito della filosofia della scienza e riconoscendogli una tradizione d'ambito anglosassone (*underdetermination*) molto più solida di quella del contesto italiano, adottiamo per definire tale caratteristica la denominazione di *sottodeterminazione* (d'ora in poi: *S.*): le potenziali alternative espressioni di *oscurità, vaghezza, indeterminatezza, ambiguità* su cui nell'ultimo trentennio si è sviluppata una pur ampia e autorevole letteratura comprendono infatti contesti in cui l'impossibilità di interpretazione letterale non è necessariamente collegata alle caratteristiche linguistiche e semantiche dell'enunciato e tendono comunque a orientarsi sugli effetti più che sulle cause del fenomeno.

Scopo del contributo è invece di presentare un paradigma teorico e metodologico di riferimento per la messa a punto di strumenti linguistico-computazionali capaci di identificare le circostanze che producono o possono produrre *S.* e quindi di contenere le possibilità di arbitraria soggettività interpretativa che la *S.* per sua stessa natura alimenta.

Le ipotesi illustrate fanno riferimento a studi condotti nelle fasi preliminari di un Progetto triennale dedicato alla normativa regionale d'ambito valdostano (§ 2).

L'attenzione qui concentrata sulla *S.* – nozione collocata alla base del costrutto di ricerca adottato – si pone l'obiettivo di evidenziarne anzitutto le specifiche e multiformi declinazioni d'ambito giuridico: imputabili a scelte in alcuni contesti necessarie e anche virtuose, ma anche - in differenti circostanze e all'opposto - a inconsapevole o intenzionale vaghezza (§ 3), le manifestazioni linguistiche della *S.* si correlano ai gradi di vincolo interpretativo posti dalle diverse tipologie testuali di genere giuridico-normativo richiedendo paradigmi descrittivi che non possono essere ricondotti tanto a parametri di *comprensibilità* quanto piuttosto a principi di *interpretabilità* e *interpretatività* (§ 4) che si ipotizza si trovino proporzionalmente distribuiti in un continuum categoriale che vede da un lato i *testi normativi* generali e astratti, *di rango primario*, e al capo opposto i testi a carattere applicativo o esecutivo. Con l'obiettivo di illustrare tale assunto viene qui testato come caso di studio l'esempio dell'uso dei lessemi appartenenti al *vocabolario di base* italiano in tre testi rappresentativi di altrettante categorie del continuum (§ 5). La prova condotta mira a supportare l'ipotesi che la *S.* non abbia (tanto o soltanto) a che fare con i vizi di eccessiva complessità o ostentazione tipicamente addebitati al linguaggio giuridico-amministrativo, ma si correli piuttosto a istanze pragmatiche – di volta in volta performative, deontiche, prescrittive, persuasive, regolative e altro ancora – che, interessando diversi livelli della lingua, si manifestano attraverso fenomeni di semantica frasale. Ai fini dell'individuazione e della puntuale descrizione di tali fenomeni la seconda fase del progetto prevede la conduzione di rilievi computazionali

su corpora bilanciati e, in parallelo, di prove sperimentali basate sulla consultazione diretta dei parlanti (§ 6): a partire dalle evidenze induttivamente ricavate e comparativamente analizzate ci si propone di giungere a delineare una fenomenologia dei marcatori linguistici assumibili come indicatori della S. giuridica. A partire da tale base nella fase conclusiva del Progetto saranno sviluppate applicazioni di supporto al *drafting* normativo (§ 7).

## 2. Il Progetto di ricerca

Condotto con un approccio interdisciplinare tra linguistica, diritto e informatica, il progetto di ricerca “La lingua del legislatore in Valle d’Aosta”<sup>4</sup> pone il proprio focus sulla normativa della piccola regione alpina che, in virtù dello Statuto di autonomia approvato con legge costituzionale 26 febbraio 1948, n.4 della Repubblica italiana, presenta proprie particolarità linguistiche e giuridiche.

Relativamente al primo aspetto – malgrado il repertorio sociolinguistico contemporaneo delinei una condizione di *bilinguismo sociale endocomunitario con dilalia multipla* (Berruto 2003: 45) in cui il francese occupa rispetto all’italiano una posizione debole in quanto a usi e competenze individuali (Raimondi 2006) – l’equiparazione statutaria<sup>5</sup> delle due lingue ufficiali ha condotto l’amministrazione regionale e gli enti locali a prevedere che entrambe e in parallelo siano utilizzate per la redazione di leggi, regolamenti, atti amministrativi, ecc.

Sulla scorta di quanto previsto dalle “Norme in materia di tutela delle minoranze linguistiche storiche” (L. 482/1999) è inoltre ammessa la possibilità di traduzione dei documenti anche nella principale parlata minoritaria locale. Proprio con l’obiettivo di “agevolare le comunicazioni in lingua francoprovenzale tra i cittadini e le istituzioni” l’Assessorato all’Istruzione e Cultura ha attivato uno specifico servizio di consulenza linguistica<sup>6</sup>.

Il potenziale plurilingue della normazione locale chiama dunque in causa tutto l’universo della trasposizione orizzontale, quindi anche insidie legate alle specificità della traduzione giuridica come le difficili equivalenze del suo lessico settoriale, le difficoltà di armonizzazione o omologazione di concetti e istituti fra contesti diversi e le esigenze di individuare soluzioni, nello specifico caso della trasposizione in francoprovenzale, per sopperire alle lacune che necessariamente una parlata locale priva di tradizione settoriale in ambito legislativo presenta.

Per ciò che concerne le particolarità dell’ordinamento giuridico, la possibilità di prevedere adattamenti locali della normativa nazionale implica poi – e soprattutto

<sup>4</sup> Il Progetto prevede un’articolazione in fasi e segmenti: al tema de “L’accessibilità dei testi normativi come dispositivo di inclusione” è dedicato uno specifico programma di attività finanziato nell’ambito del bando 2021 “Erogazioni ordinarie” della Fondazione CRT.

<sup>5</sup> Art. 38, commi 1 e 2: “Nella Valle d’Aosta la lingua francese è parificata a quella italiana. Gli atti pubblici possono essere redatti nell’una o nell’altra lingua, eccettuati i provvedimenti dell’autorità giudiziaria, i quali sono redatti in lingua italiana”.

<sup>6</sup> Cfr. [www.patoisvda.org](http://www.patoisvda.org).



in caso di *S.* del dettato normativo statale – che le possibilità interpretative a disposizione del legislatore regionale determinino esiti inattesi o anche conflittuali rispetto alle competenze dello Stato.

Con queste premesse, il progetto avviato presso l'Università della Valle d'Aosta mira a definire circostanze e fenomenologia della *S.* esaminandone le declinazioni all'interno delle diverse dimensioni implicate nel contesto valdostano assunto a caso di studio: a livello intralinguistico, in rapporto alla riformulazione del dettato statale per l'introduzione degli adattamenti locali; a livello interlinguistico, in relazione alla trasposizione dei testi normativi regionali nella seconda lingua di cultura e potenzialmente anche nella parlata minoritaria endogena; nella dimensione produttiva, con riguardo alle diverse fasi del *drafting* normativo e dei processi di redazione di conseguenti atti applicativi; in quella ricettiva, in collegamento alle dimensioni linguistiche e giuridiche dei processi interpretativi.

### 3. *Sottodeterminazione e universi giuridici*

Le questioni legate alla 'buona scrittura delle leggi' sono state esaminate, soprattutto nell'ultimo ventennio e anche in sedi istituzionali, attraverso approcci linguistici e giuridici integrati. All'interno dei dibattiti scaturiti dal confronto interdisciplinare il problema dell'oscurità espressiva è stato ricondotto dai linguisti principalmente a questioni di leggibilità e comprensibilità testuale e inquadrato secondo prospettive orientate sui destinatari nell'ambito dei contesti della comunicazione pubblica; è stato d'altra parte riconosciuto dai giuristi come radicato vezzo legato ad ataviche istanze di salvaguardia del valore magico della lingua giuridica (Ondelli 2007), a intenti criptolalic conservativi dei privilegi di cerchie ristrette, a cifre stilistiche di ostentazione erudita, ma anche descritto come in alcuni contesti indispensabile dispositivo di garanzia della flessibilità e longevità di norme altrimenti soggette a rapida obsolescenza.

Casi di *S.* possono ad esempio essere identificati nella *Costituzione italiana*, il cui testo – coralmemente elogiato per efficacia di formulazione – lascia in diversi luoghi possibilità di spazi interpretativi alternativi. Citiamo, a titolo di esempio, il comma 2 dell'art. 59, che nella versione originale recitava:

Il Presidente della Repubblica può nominare senatori a vita cinque cittadini che hanno illustrato la Patria per altissimi meriti nel campo sociale, scientifico, artistico e letterario.

L'oscillazione fra interpretazioni restrittive o estensive, a seconda che per *Presidente della Repubblica* si intendesse l'organo costituzionale o invece l'individuo che ricopre l'incarico, ha inciso in alcune circostanze della storia italiana recente sulla sopravvivenza di maggioranze politiche per le quali il voto dei senatori a vita risultava decisivo. Una recente revisione costituzionale (art. 3 comma 1 della L.C. n. 1 del 19 ottobre 2020) ha infine risolto la *S.* della formulazione originaria attraverso la seguente integrazione destinata a definitiva disambiguazione: "Il numero com-

pllessivo dei senatori in carica nominati dal Presidente della Repubblica non può in alcun caso essere superiore a cinque”.

La possibilità d'intervenire sul testo dopo aver sperimentato i concreti effetti di formulazioni deliberatamente sottodeterminate consente di evitare decisioni aprioristiche, potenzialmente foriere di esiti imprevedibili: risulta d'altra parte chiaro che gli usi della *S.* ispirati a tale principio di flessibilità o cautela possano trovare maggiore o minore legittimazione a seconda delle categorie testuali in cui si collocano e delle cause che ne motivano la presenza.

Partendo dall'inquadramento proposto in Longo, Mastropaolo e Pallante (2012) e approfondendone la fenomenologia, proponiamo allora anzitutto di distinguere i fenomeni di *S.* deliberatamente adottata dal legislatore (*S. fisiologica*) da quelli in cui la *S.* non svolge invece funzioni giuridicamente motivate (*S. patologica*).

Alla categoria della *S. fisiologica* appartengono i casi in cui il legislatore con intenzionalità seleziona fra le molte disponibili le soluzioni linguistiche capaci di garantire una comprensività dell'enunciazione indispensabile in presenza di una pluralità di possibili scenari applicativi, presenti o futuri, così come quelli in cui un certo grado d'indeterminatezza espressiva rappresenta un'ineludibile mediazione strategica fra istanze politiche differenti e altrimenti inconciliabili. La presenza di *S. fisiologica* in circostanze di questo tipo non inficia la normatività della *produzione legislativa*: semplicemente rimanda la puntualità di concrete declinazioni di significato attraverso una delega interpretativa.

Costituiscono invece esempi di *S. patologica* i casi di equivocità degli enunciati determinati da insufficiente elaborazione concettuale del legislatore; le ambiguità conseguenti a sciatteria linguistica e scrittoria; le formulazioni criptiche introdotte per l'esibizione di virtuosismi stilistici. In questi casi la presenza di *S.* corrisponde a tutti gli effetti a *cattiva produzione legislativa* ed espone gli enunciati a un' indesiderabile arbitrarietà interpretativa.

Da tale assunto discendono due corollari. Il primo prevede che laddove la *S. fisiologica* sia sempre controllata e intenzionalmente introdotta dal legislatore, la *S. patologica* non risponda invece a una reale necessità e possa quindi essere quanto meno circoscritta attraverso riformulazioni linguistiche senza che questo comporti sacrifici o distorsioni del contenuto regolativo. Il secondo fa riferimento alla disposizione delle categorie testuali della letteratura d'ambito giuridico all'interno di un continuum ai cui estremi si collocano rispettivamente i *testi normativi di rango primario*, generali e astratti secondo la classica definizione della legge, e all'opposto i *testi normativi di rango secondario*, applicativi o esecutivi, con funzioni illocutive e carattere pragmatico: in quest'ottica sembra possibile ipotizzare che presentino maggiori necessità e caratteristiche di *S. fisiologica* i testi di rango primario, per i quali si prevedono ampio respiro, caratteristiche di generalità e una quanto più ampia possibile longevità; che debbano sottrarsi alla *S. patologica* – richiedendo invece la massima coerenza e univocità interpretativa – le tipologie testuali di prospettiva applicativa, di breve prospettiva temporale e/o limitato ambito di applicazione.

Anche con riguardo ai pubblici di riferimento e rispetto all'asse dei rapporti fra *S.* e comprensibilità sembra d'altra parte opportuno distinguere requisiti e caratteristiche dei testi e degli enunciati a carattere prettamente applicativo-amministrativo da quelli più propriamente normativo-giuridici, giacché, come avverte Rovere (2007: 202), «la comprensibilità di un testo giuridico per un interlocutore profano non implica necessariamente che per il giurista il testo sia univoco. Viceversa, rendere comprensibili testi giuridici a non specialisti significherebbe mettere a loro disposizione, in aggiunta al testo, tutto il sapere giuridico indispensabile a tal fine, oppure accettare il rischio, altamente probabile, di imprecisioni e ambiguità tecniche».

#### 4. *Sottodeterminazione e grado di vincolo interpretativo: interpretabilità e interpretatività*

Proposto e a più riprese approfondito da Francesco Sabatini, anche con espresso riferimento all'ambito normativo, il modello di classificazione delle tipologie testuali basato sul *grado di vincolo interpretativo* attribuisce ai testi giuridici la natura di scritture collocate al 'polo della massima rigidità ed esplicitezza' (2011b: 338), con *discorso molto vincolante* e "funzione prescrittiva, basata su una manifestazione di volontà coercitiva, regolata da un intero sistema di principi enunciati espressamente". Precisa Sabatini (2011a: 192) che l'intenzione dell'emittente "di regolare in maniera più o meno rigida l'attività interpretativa del destinatario" presenta due componenti, che bene si prestano a delineare una prima definizione delle circostanze della *S.*

La prima, descritta come "atteggiamento di personale e generica attenzione dell'emittente nel cercare di rendere ricevibile il senso complessivo del messaggio" tratteggia in riferimento a un enunciato giuridico la sua *interpretabilità* intesa come concreta possibilità decifrativa del significato globale. All'interno del patto comunicativo tra emittente e destinatario, le possibilità che il testo risulti decodificabile (*ricevibile*) hanno necessariamente a che fare con le specifiche competenze del ricevente: per garantire a un pubblico quanto più ampio possibile l'accessibilità di un enunciato giuridico occorre che questo sia formulato nel rispetto degli indicatori canonici della comprensibilità e leggibilità.

La seconda componente, che attribuisce all'estensore del testo un "comportamento di forte consapevolezza delle modalità di interpretazione puntuale di ogni elemento linguistico del suo messaggio", richiama invece condizioni di possibilità esegetica che non possono essere ricondotte a parametri linguistici di superficie ma evocano piuttosto dispositivi ermeneutici a base semantica. Ci pare sia questa seconda condizione, che denominiamo *interpretatività*, a chiamare in causa la *cogenza* o al contrario la *S.* di un enunciato o di un testo giuridico.

Poniamo allora alla base del nostro paradigma di analisi una prima fondamentale distinzione tra i concetti di *interpretabilità* e di *interpretatività*: le conseguenti ipotesi riguardano la possibilità che un enunciato giuridico caratterizzato da *interpretabilità* non necessariamente presenti anche caratteristiche di *interpretatività*; che in

alcuni casi, anzi, *interpretabilità* e *interpretatività* siano inversamente proporzionali, come avviene quando con l'obiettivo di "rendere ricevibile il senso complessivo del messaggio" le possibilità di "interpretazione puntuale di ogni elemento linguistico" possano essere sfavorite da una riduzione dell'intensione degli enunciati e quindi dal ricorso, intenzionale o accidentale, alla *S*.

Ipotizziamo, inoltre, che la *S*. si corredi all'*interpretatività* concentrandosi nella sua veste *fisiologica* – all'interno del continuum delle categorie testuali ascrivibili all'ambito giuridico – nelle formulazioni dei testi normativi di rango primario, gradatamente scomparendo o rarefacendosi nei testi applicativi o esecutivi.

L'ovvia constatazione che questo nella realtà concreta non sempre accade, piuttosto che a mettere in discussione l'ipotesi teorica, ci induce a porci un interrogativo: e se in alcuni casi fosse l'intenzione del legislatore di favorire l'*interpretabilità* ad amplificare gli spazi d'*interpretatività* di un testo giuridico?

Cerchiamo di contestualizzare la domanda assumendo come caso esemplificativo uno dei fondamenti della *scrittura controllata*, che compare costantemente nelle raccomandazioni contenute nelle linee guida per la redazione di testi chiari e comprensibili: l'indicazione a un uso quanto più possibile esteso di parole della lingua comune. Il punto di riferimento usualmente assunto per l'identificazione di tali parole all'interno del serbatoio lessicale dell'italiano contemporaneo è quello del *Vocabolario di Base* (VdB) pubblicato per la prima volta da Tullio De Mauro come appendice alla sua *Guida all'uso delle parole* (1980) e riedito nel 2016 in una nuova versione basata su un più aggiornato spoglio e campionamento di testi (*Nuovo Vocabolario di Base: NVdB*).

Come noto, entrambe le versioni del repertorio lessicale comprendono circa settemila vocaboli ripartiti in tre categorie: le prime due, elaborate sulla base dell'analisi statistica di ampi campioni di testi appartenenti a differenti tipologie, sono quelle del *lessico fondamentale* (FO: le circa duemila parole di maggior uso assoluto) e del *lessico di alto uso* (AU: circa duemilasettecento termini ad altissima frequenza); la terza, quella del cosiddetto *lessico di alta disponibilità* (AD), raccoglie invece circa duemilatrecento vocaboli che sulla base di indagini sul campo risultano essere percepiti dai parlanti comuni come molto familiari pur non presentando frequenze statistiche di rilievo.

Il parametro dell'appartenenza o non appartenenza dei lessemi al *vocabolario di base* e alle sue categorie interne è stato assunto nell'ultimo quarantennio come indicatore per la valutazione della leggibilità dei testi in relazione a diversi ambiti applicativi della scrittura controllata e anche attraverso strumenti prodotti nell'ambito della linguistica computazionale<sup>7</sup>. In linea generale, scarsa considerazione è però stata accordata a quei potenziali effetti collaterali della frequenza lessicale da cui lo stesso De Mauro (2009) aveva messo in guardia osservando che "data una lista

<sup>7</sup> Più recentemente l'Istituto di Linguistica Computazionale "Antonio Zampolli" del CNR di Pisa ha messo a punto uno strumento di valutazione avanzata della leggibilità – READ-IT – che identifica la percentuale di lemmi del vocabolario del testo non appartenenti al Vocabolario di Base e per quelli ad esso riconducibili assume a riferimento le relative classiche tre categorie (Dell'Orletta et al. 2011).

di frequenza o uso delle parole di una qualsiasi lingua, i lessemi più frequenti hanno in media un maggior numero di accezioni, più esattamente al crescere della frequenza secondo potenze cresce secondo la successione naturale il numero delle accezioni". Pare probabile che un così significativo fenomeno di proliferazione semantica possa assumere contorni molto specifici nell'ambito delle lingue speciali, che per loro natura tendono a convertire i termini del lessico comune verso sensi tecnici o specializzati e comunque non corrispondenti a quelli ordinariamente impiegati e noti ai parlanti. Il parametro lessicometrico dell'alta frequenza potrebbe allora in ambito giuridico non garantire di per sé l'accesso ai corretti significati contestuali e anzi alimentare, come in alcune altre varietà dell'italiano (Revelli 2020), ingannevoli rappresentazioni di trasparenza e comprensibilità: il ricorso a termini di solo apparente facilità interpretativa incrementerebbe, invece che arginare, le circostanze della *S.* di un enunciato giuridico.

### 5. Sottodeterminazione e comprensibilità: l'esempio delle frequenze lessicali

Approfondiamo la questione verificando, in primo luogo, come i livelli di polisemia – misurati sulla base di *MultiWordNet* (MWN nel seguito; Pianta et al. 2002; Toral et al. 2010), un *dump* italiano della risorsa lessicografica WordNet (Miller 1995), – si distribuiscano nelle due versioni del *Vocabolario di base* (versione 1980<sup>8</sup>: VdB; versione 2016<sup>9</sup>: NVdB) all'interno delle tre categorie di frequenza lessicale demauriane.

Come illustrato in tab. 1 e in tab. 2, sia nel VdB che nel NVdB i lessemi etichettati come FO appaiono significativamente più polisemici di quelli marcati AU, a loro volta più polisemici di quelli ad AD. Le evidenze relative al tasso di lemmi polisemici concordano sostanzialmente con le indagini già condotte in precedenza sul livello di polisemia del VdB (Casadei 2014).

Nell'indagine sul livello di polisemia caratterizzante le tre partizioni del VdB, i due indicatori ('% polysem' e 'Syn/Lem' nelle tab. 1 e 2) possono essere posti in relazione: il primo misura la proporzione di elementi polisemici (quanti sono i lemmi polisemici: frequenza relativa), mentre il secondo coglie la consistenza di detta polisemia (quanto in media sono polisemici i lemmi di una data categoria), essendo calcolato come rapporto fra numero di *synset* (insiemi di sinonimi in *WordNet*) e numero di lemmi<sup>10</sup>. Le analisi condotte illustrano quindi non solo come all'interno dei lemmi FO siano più frequenti elementi polisemici (rispetto ad AU e AD: para-

<sup>8</sup> In De Mauro 1980: 149-183.

<sup>9</sup> [www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana](http://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana).

<sup>10</sup> Va notato come le percentuali di termini polisemici debbano essere integrate da quella di termini privi di *synset* associati in *MWN*. La frazione di termini non presenti nella base lessicografica cresce passando dal vocabolario ad AU, e da questo a quello ad AD. Tale andamento è comune alle due versioni del VdB, ed è compatibile con l'ipotesi che lemmi meno frequenti siano non solo meno presenti all'interno della risorsa lessicografica, ma anche connotati da un minore grado di polisemia.

metro “% polysem”), ma anche come nel passaggio da FO, AU e AD tale polisemia sia caratterizzata da un numero di sensi decrescente (parametro “Syn/Lem”).

Tabella 1 - Livello di polisemia nelle tre sezioni del VdB, misurata come numero medio di *synset* per lemma, percentuale di lemmi non presenti in MWN, percentuale di termini monosemici e polisemici

VdB	N. Lemmi	Syn/Lem	% no <i>synsets</i>	% monosem	% polysem
FO	1855	4.1	5.12%	13.05%	81.83%
AU	2922	2.3	10.40%	30.18%	59.41%
AD	1632	1.7	12.38%	41.12%	46.51%

Tabella 2 - Livello di polisemia nelle tre sezioni del NVdB, misurata come numero medio di *synset* per lemma, percentuale di lemmi non presenti in MWN, percentuale di termini polisemici

NVdB	N. Lemmi	Syn/Lem	% no <i>synsets</i>	% monosem	% polysem
FO	1979	4.1	5.36%	13.14%	81.51%
AU	3011	2.2	9.66%	31.12%	59.22%
AD	2231	1.5	17.71%	43.30%	39.00%

Verifichiamo allora se, come ipotizzato, l’esigenza di circostanziare in modo univoco e vincolante l’interpretazione del testo conduca il legislatore a identificare nel VdB, quindi nel lessico più polisemico, le maggiori insidie della *S*. Ci aspettiamo quindi che il ricorso a termini del VdB sia maggiore nei testi di rango primario, che, come evidenziato in § 4, richiedono una comprensione globale e tollerano o anzi implicano un certo livello di *S*.; che sia invece minimo nei testi regolativi, le cui funzioni illocutive rispondono all’esigenza pragmatica di fornire prescrizioni inequivocabili.

Conduciamo un test preliminare considerando tre documenti paradigmatici dei due poli dell’ideale continuum delle tipologie testuali d’ambito giuridico – la *Costituzione italiana* e il *Decreto Asseverazioni 6 agosto 2020*<sup>11</sup> ai due poli opposti, e la *Legge 240/2010*<sup>12</sup> in una posizione centrale, rappresentativa delle tipologie giuridiche miste di rango intermedio.

I dati rappresentati nella tab. 3 consentono di osservare come nei tre testi oggetto di analisi il livello di polisemia si riduca parallelamente alla riduzione della copertura dei termini del VdB. Tale riscontro può essere integrato dall’andamento dei valori nella prima colonna della tabella, che riporta la percentuale di termini non associati ad alcun *synset*. Questi sono meno numerosi nella *Costituzione*, mentre crescono significativamente per il testo del *Decreto*, rappresentativo dei testi di rango secondario. Al di là di questioni connesse alla copertura della risorsa lessicale

<sup>11</sup> Si tratta del Decreto del Ministero dello Sviluppo economico 6 agosto 2020 recante “*Requisiti tecnici per l’accesso alle detrazioni fiscali per la riqualificazione energetica degli edifici – cd. Ecobonus*” (GU n. 246 del 05-10-2020, Serie Generale).

<sup>12</sup> L. n. 240 del 30 dicembre 2010: “*Norme in materia di organizzazione delle università, di personale accademico e reclutamento, nonché delega al Governo per incentivare la qualità e l’efficienza del sistema universitario*” (GU n. 10 del 14.1.2011, Suppl. Ordinario n. 11).



utilizzata (Pianta et al. 2002), un simile andamento è compatibile con il ricorso a un lessico progressivamente più specifico passando dalla *Costituzione*, attraverso la *L. 240/2010*, fino al *Decreto*.

Tabella 3 - *Livello di polisemia dei tre documenti assunti a campione: le prime tre colonne illustrano rispettivamente la percentuale di termini non presenti in MWN, quelli associati a un solo synset e quelli polisemici; le colonne successive illustrano invece il numero di termini presenti all'interno delle tre sezioni del VdB e la copertura totale, espressa come somma*

VdB	% no synsets	% monosem	% polysem	cover. FO	cover. AU	cover. AD	$\Sigma$ cov.
Costituz. ITA	16.37%	16.01%	67.62%	48.37%	17.84%	2.44%	68.65%
L. 240/2010	17.25%	19.39%	63.36%	38.44%	15.73%	1.96%	56.13%
D. 6.08.2020	25.82%	18.83%	55.35%	21.21%	6.88%	1.24%	29.33%

Per quanto riguarda la percentuale dei termini monosemici i valori rilevati non consentono di individuare tendenze intelligibili (per ampiezza e direzione della variazione), mentre il livello di polisemia è decrescente nel passaggio dalla *Costituzione* al *Decreto*. Il livello della correlazione con la copertura dei tre documenti del VdB è complicato dall'interazione di vari fattori, quali la completezza della risorsa utilizzata, ma anche da altri elementi rilevanti seppur di natura implementativa<sup>13</sup>, e richiede ulteriori analisi ed approfondimenti.

Sembra plausibile l'ipotesi che nei testi normativi di rango primario l'ampio ricorso a termini del VdB risponda al contempo a un obiettivo di raggiungibilità di un pubblico ampio e non specialista e a una necessità di *S.* volta a garantire ai contenuti del dettato l'opportuna longevità e flessibilità. Nei testi di rango secondario, un uso parsimonioso del VdB associato al ricorso a tecnicismi sarebbe invece rivolto a evitare una *S.* che, se presente, comporterebbe necessariamente ambiguità applicative.

La ricognizione qui proposta ha sondato unicamente il livello della polisemia di unità lessicali monorematiche con l'obiettivo di verificare preliminarmente e su un ambito circoscritto l'ipotesi che esista una gerarchia dell'ambiguità lessicale parallela alla *S.* rappresentata nel continuum tipologico delle fonti giuridiche. Si ipotizza che la *S.* che caratterizza le diverse categorie di documento si associ tuttavia non soltanto a ulteriori dimensioni semantico-lessicali qui non considerate e a tratti specifici del linguaggio giuridico-normativo, ma anche alle loro interfacce con più profondi livelli della lingua. Con lo scopo di verificare questa ipotesi e di individuare indicatori che possano conclusivamente essere assunti come marche linguistiche identificative della *S.* giuridica, la fase sperimentale del Progetto – che prevede la costituzione ad hoc di un corpus normativo bilanciato da sottoporre a trattamento computazionale e la parallela conduzione di indagini sul campo – seguirà le coordinate sinteticamente qui di seguito descritte.

<sup>13</sup> Come, per esempio, il filtraggio delle *stopwords*, per cui si è inizialmente utilizzato un insieme di *parole piene* messe a disposizione dalla risorsa NLTK (Loper & Bird 2002), a una prima analisi largamente incompleto e non adattato al contesto applicativo corrente.

### 6. Per l'individuazione di indicatori linguistici ad hoc

Posto che quella di *linguaggio giuridico* è una “etichetta di estensione variabile, adattata [...] a un universo testuale composito, in cui si riconoscono varietà di lingua concorrenti” (Mortara Garavelli 2001: 7) e che all'interno della letteratura del settore normativo tratti microlinguistici possono assumere funzionalità tecniche e stilistiche differenziate, gli indicatori della *S.* da identificare devono necessariamente tenere conto delle interfacce che possono interagire ai diversi livelli della lingua e contraddistinguere gli specifici contesti pragmatici, oltre che tipologico-testuali.

Per selezionare fra i tratti descritti in letteratura come caratteristici del linguaggio giuridico solo quelli effettivamente significativi in relazione alla *S.* e con lo scopo di farne emergere di ulteriori, eventualmente condivisi con le microlingue di altri ambiti settoriali o con differenti generi ad alta vincolatività, sono state previste due classi di procedure sperimentali.

La prima prevede che a partire da un censimento di leggi e regolamenti prodotti in Valle d'Aosta dal 1950 ai giorni nostri si giunga a definire una tassonomia delle tipologie testuali giuridiche locali secondo parametri riconducibili a una scala gerarchica che permetta di declinare con un soddisfacente livello di dettaglio la tipologia dei generi rappresentati. La selezione di alcune delle tipologie individuate come maggiormente caratteristiche della realtà esaminata sarà poi utilizzata per la costituzione di un corpus testuale qualitativamente e quantitativamente rappresentativo di diverse prassi redazionali. All'interno di tale corpus si procederà alla creazione di *sottocorpora* distinti per la costituzione di concordanze e lessici di frequenza. Test di leggibilità testuale consentiranno di verificare se i luoghi di maggiore criticità identificati secondo gli indicatori canonicamente impiegati in letteratura consentono di rilevare manifestazioni della *S. fisiologica e patologica*, ed in rapporto a quali specificità linguistiche. I documenti saranno poi sottoposti ad analisi computazionali comparative volte a ricavare indici di ricchezza lessicale e densità semantica utili anche per l'interpretazione dell'andamento delle frequenze lessicali e per la parametrizzazione di indicatori relativi ai rapporti fra sinonimia, monosemia e polisemia, iponimia e iperonimia. In questa fase sarà esplorata inoltre la *S.* in analoghi testi in altre lingue con rappresentazioni vettoriali come *word-* e *sense-embeddings* multilinguistici (Colla et al. 2020) e contestuali (Devlin et al. 2018), per verificare l'ipotesi dell'utilizzo graduato della *S.* in corrispondenza con i diversi livelli cui appartengono i documenti giuridici.

La seconda classe di procedure prevede la conduzione di prove sul campo con il coinvolgimento di tre campioni di parlanti: il primo gruppo comprenderà i *produttori* di testi giuridici, ossia addetti ai lavori direttamente coinvolti nelle diverse fasi di stesura delle norme (livello politico); il secondo gruppo sarà costituito da funzionari amministrativi con ruoli di interpreti esecutivi dei contenuti normativi (livello tecnico); il terzo gruppo sarà formato da parlanti comuni.

Alle tre coorti sarà proposta, in tre distinte fasi, la lettura di testi normativi rappresentativi di categorie tipologiche a vincolatività differenziata.



La prima fase prevedrà la lettura a voce alta di un testo proposto sul monitor di un computer: attraverso un'applicazione per l'*eye-tracking* sviluppata *ad hoc* saranno registrati pattern di movimenti oculari e rilevati movimenti anomali, quali il protrarsi di fissazioni su specifici punti, le variazioni di ritmo e estensione dei movimenti saccadici progressivi; la comparsa e l'intensificazione di movimenti saccadici regressivi (ritorni all'indietro orizzontali, da destra verso sinistra, e verticali, dal basso verso l'alto). Ci si attende che l'analisi dei tracciamenti condotta congiuntamente all'osservazione dei modi e tempi di esecuzione del compito di lettura consenta di identificare le variazioni attenzionali associate a difficoltà interpretative e quindi a mettere in relazione le collocazioni rilevate con l'analisi dei risultati dell'analisi dell'*eye tracking* (*S. tracciata*) con le circostanze linguistiche che le caratterizzano.

La seconda fase consisterà in una rilevazione off line della *S. percepita* dai lettori. Tramite domande poste immediatamente dopo la lettura dei testi attraverso un questionario semistrutturato ci si proporrà di verificare da un lato la comprensione di specifici passi e l'interpretazione di proposizioni potenzialmente ambigue; d'altro lato si chiederà agli intervistati di esprimere le proprie opinioni rispetto alla chiarezza complessiva del documento con valutazioni quantitative espresse su scala Likert e domande a risposta aperta. In tal modo si mira a ricavare evidenze comparative sui livelli di comprensione riscontrabili all'interno delle tre categorie di parlanti considerate e al contempo a ottenere informazioni circostanziate sulle loro rappresentazioni della *S. giuridica*.

Nella terza fase ci si proporrà di rilevare le modalità della *S. processata* attraverso operazioni di traduzione di un testo giuridico dall'italiano verso il francese. Per la conduzione della prova sarà utilizzata la tecnica del protocollo del pensiero ad alta voce: si chiederà quindi ai soggetti intervistati di procedere alla trasposizione avvalendosi al bisogno di un dizionario bilingue ed esplicitando di volta in volta dubbi e criticità, in modo che le difficoltà legate a *S.* del testo di partenza possano in tempo reale, durante lo svolgimento del compito, essere rilevate, registrate e distinte da quelle imputabili a ragioni diverse.

A conclusione delle sperimentazioni, la comparazione delle evidenze risultanti dal trattamento computazionale del corpus con gli esiti delle prove sul campo dovrà condurre a identificare gli indicatori di *interpretatività* distinguendoli, ove necessario, da quelli impiegati per la valutazione dell'*interpretabilità*, anche in relazione a singole specificità della variazione semasiologica dei sottocodici giuridici. Sarà quindi predisposto un inventario dei tratti linguistici identificabili come marcatori della *S.*: un'analisi dei contesti di occorrenza si proporrà di classificarli da un punto di vista tipologico – semantico e onomasiologico – per la strutturazione di un sistema di etichettatura rappresentativo della casistica censita.

### 7. Sviluppi applicativi: sistemi di supporto linguistico-computazionale

La predisposizione di applicazioni informatiche concepite ad hoc consentirà di verificare presenza e distribuzione delle differenti etichette tipologiche della *S.* nei

diversi generi testuali rappresentati nel corpus. La verifica della coerenza dei riscontri, testata anche al di fuori degli ambiti documentari compresi nel campione di partenza, dovrà infine condurre all'introduzione delle modifiche eventualmente necessarie per la validazione conclusiva sia del modello categoriale delle tipologie testuali d'ambito giuridico sia dell'inventario degli indicatori linguistico-giuridici parametrati.

I successivi sviluppi applicativi potranno essere rivolti alla predisposizione di linee guida e alla costruzione di un sistema di supporto a vari compiti, che nelle fasi di *drafting* normativo possa favorire – a seconda delle necessità legate ai diversi contesti testuali – il ricorso consapevole e mirato alla *S.* o il suo opportuno intenzionale evitamento. In dimensione ricettiva, tale sistema dovrà agevolare il riconoscimento delle circostanze della *S.*, favorendone l'interpretazione sia in direzione applicativa, sia nella prospettiva traduttiva, a livello intralinguistico (trasposizione verticale a fini divulgativi) e interlinguistico (*task* della traduzione propriamente intesa).

In quest'ottica ci si auspica, conclusivamente, che i modelli teorici e gli strumenti elaborati attraverso il paradigma della *S.* possano andare a integrare in modo complementare i contributi già nell'ultimo decennio apportati dalle *digital humanities* al generale miglioramento della qualità degli atti normativi<sup>14</sup>.

### Bibliografia

- Berruto, Gaetano. 2003. Una Valle d'Aosta, tante Valli d'Aosta? Considerazioni sulle dimensioni del plurilinguismo in una comunità regionale. In Fondation Emile Chanoux, *Una Valle d'Aosta bilingue in un'Europa plurilingue*, 44-53. Aosta: Tipografia Valdostana.
- Casadei, Federica. 2014. La polisemia nel vocabolario di base dell'Italiano. *Lingue e Linguaggi* 12. 35-52.
- Colla, Davide & Mensa, Enrico & Daniele P. Radicioni. 2020. LessLex: Linking multilingual Embeddings to SenSe representations of lexical items. *Computational Linguistics*. 46(2). 289-333.
- De Mauro, Tullio. 1980. *Guida all'uso delle parole*, 149-183. Roma: Editori Riuniti.
- Dell'Orletta, Felice & Montemagni, Simonetta & Venturi, Giulia. 2011. READ-IT: Assessing readability of Italian texts with a view to text simplification. In *SLPAT '11. Proceedings of the second Workshop on Speech and Language Processing for Assistive Technologies*, 73-83. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Devlin, Jacob & Chang, Ming-Wei & Lee, Kenton & Toutanova, Kristina. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint 1810.04805.

<sup>14</sup> In particolare, la famiglia dei prodotti *xmLeges*, sviluppati dall'Istituto di Teoria e Tecnica dell'Informazione Giuridica nell'ambito del Progetto *Norme in Rete* con l'obiettivo di supportare la redazione, revisione e conversione in XML di testi normativi redatti in conformità alle regole di *drafting* legislativo e agli standard di formato previsti per la rappresentazione elettronica dei provvedimenti (Rinaldi 2016).

- Longo, Fabio & Mastropaolo, Antonio & Pallante, Francesco. 2012. Incertezze derivanti dalla ineliminabile, ma non adeguatamente contenuta, oscurità linguistica delle disposizioni normative. In Dogliani, Mario (a cura di), *Il libro delle leggi strapazzato e la sua manutenzione*, 43-49. Torino: Giappichelli.
- Loper, Edward & Bird, Steven. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.
- Miller, George A. 1995. WordNet: A lexical database for English. *Communications of the ACM*. 38(11). 39-41.
- Mortara Garavelli, Bice. 2001. *Le parole e la giustizia. Divagazioni grammaticali e retoriche su testi giuridici italiani*. Torino: Einaudi.
- Ondelli, Stefano. 2007. *La lingua del diritto: proposta di classificazione di una varietà dell'italiano*. Aracne: Roma.
- Pianta, Emanuele & Bentivogli, Luisa & Girardi, Christian. 2002. MultiWordNet: Developing an aligned multilingual database. In *Proceedings of the first International Conference on Global WordNet*, 293-302. Mysore, India.
- Raimondi, Gianmario. 2006. Storia e configurazione del repertorio plurilingue valdostano. In Bertolino, Fabrizio & Revelli, Luisa (a cura di), *Università, scuola, territorio*, 100-126. Milano: Franco Angeli.
- Raso, Tommaso. 2016 (2005). *La scrittura burocratica. La lingua e l'organizzazione del testo*. Roma: Carocci.
- Revelli, Luisa. 2020. Lexicometric paradoxes of frequency: Comparing VoBIS and NVdB. In Iezzi, Domenica & Mayaffre, Damon & Misuraca, Michelangelo (eds.), *Text analytics: Advances and challenges*, 91-99. Cham: Springer.
- Rinaldi, Mariasole. 2016. XmLeges-Editor: un ambiente completo per la redazione e conversione di testi normativi. In Panizza, Saulle (a cura di), *Profili attuali di qualità degli atti normativi e amministrativi*, 247-268. Pisa: University Press.
- Rovere, Giovanni. 2007. L'efficienza comunicativa in ambito giuridico. *Annali del Dipartimento di Scienze Giuridico-Sociali. Università del Molise* 9. 183-208.
- Sabatini, Francesco. 2011a (1ª ed. 1999). "Rigidità-esplicitzza" vs "elasticità-implicitzza": possibili parametri massimi per una tipologia dei testi. In Sabatini, Francesco, *L'italiano nel mondo moderno. Saggi scelti dal 1968 al 2009*, tomo II, 183-216. Napoli: Liguori.
- Sabatini, Francesco. 2011b (1ª ed. 2001). I tipi di testo e la "rigidità" del testo normativo giuridico. In Sabatini, Francesco, *L'italiano nel mondo moderno. Saggi scelti dal 1968 al 2009*, tomo II, 335-346. Napoli: Liguori.
- Toral, Antonio & Bracale, Stefania & Monachini, Monica & Soria, Claudia. 2010. Rejuvenating the Italian WordNet: Upgrading, standarisising, extending. In *Proceedings of the 5th Global WordNet Conference (GWC)*, Bombay, India: ACL.