



Istituto di Teoria e Tecniche dell'Informazione Giuridica
Consiglio Nazionale delle Ricerche

Note sulla implementazione di regole formali
per il riconoscimento automatico dei riferimenti
normativi espliciti, delle novelle, delle
disposizioni di delega

a cura di Pietro Mercatali e Francesco Romano
Firenze - 2003
Rapporto tecnico n. 1/2003

NOTE SULLA IMPLEMENTAZIONE DI REGOLE FORMALI PER IL RICONOSCIMENTO AUTOMATICO DEI RIFERIMENTI NORMATIVI ESPLICITI, DELLE NOVELLE, DELLE DISPOSIZIONI DI DELEGA

1. Finalità e presupposti

La farraginosità del sistema giuridico è un ostacolo, tutt'altro che secondario, all'evoluzione di molte società, nelle quali si sono dilatati a dismisura i tempi (e quindi i costi) per l'ottenimento di una qualsiasi decisione rilevante per il cittadino, da parte dei sistemi giudiziario e amministrativo.

Presupposto a qualsiasi proposta e intervento per il superamento di tale immobilismo è l'accessibilità e la circolazione dei dati pubblici, in particolare di quelli normativi.

Sono note le molteplici attività e le ingenti risorse investite dalle politiche di e-government di molti paesi, per rendere le tecnologie informatiche e telematiche un potente strumento al servizio di tali obiettivi.

L'esigenza propedeutica a tale sforzo è di disporre di linguaggi standard, che le macchine siano in grado di "comprendere" e trattare. Al tempo stesso è però indispensabile conciliare l'impiego di tali linguaggi per il dialogo uomo-macchina, con l'accesso e la fruizione dell'informazione da parte dei cittadini. L'accesso e la fruizione di massa dei dati pubblici, pur veicolati dalle macchine, non possono realizzarsi con linguaggi artificiali, ma solo con formati comunicativi non specialistici come il linguaggio naturale.

Il trattamento e il riconoscimento automatico del linguaggio naturale, appare assumere un ruolo fondamentale in tale interazione uomo-macchina-uomo.

Per la comunicazione delle fonti normative tramite Internet, le istituzioni parlamentari e governative di molti paesi hanno avviato un processo di conversione dei loro "giacimenti" in un formato standard per facilitare la ricerca e visualizzazione dei testi.

Il linguaggio di marcatura XML sembra essere lo strumento deputato a raggiungere lo scopo. Tale linguaggio, infatti, coniugando la sua duplice natura di linguaggio di marcatura e di standard WEB, può costituire il terreno comune sia degli interventi "a monte", cioè sul drafting legislativo, sia di quelli "a valle" che concernono la pubblicazione dei testi e l'individuazione di strumenti per l'accesso all'informazione legislativa.

Per l'adozione di tale linguaggio come standard e, soprattutto, per la conversione dei testi normativi vigenti in tale formato è necessaria, a nostro avviso, l'interazione di due fattori.

A) Definizione e promozione di un linguaggio normativo "controllato".

La legistica o tecnica legislativa ha introdotto elementi univoci e ricorrenti nei testi normativi, per cui si può identificare nel linguaggio normativo un linguaggio più controllato rispetto a quello comune. Si sono infatti adottate specifiche norme di ortografia, lessico, sintassi, stile e struttura per la stesura dei testi legislativi. Una raccolta di tali regole è contenuta in una circolare del 2001 emanata dai Presidenti del Consiglio, Camera e Senato ed adottata da Governo e Parlamento. La circolare ne aggiorna una precedente del 1986. Pressoché tutte le regioni, per la redazione dei loro atti normativi, hanno adottato il Manuale "Regole e suggerimenti per la redazione dei testi normativi" compilato nel 1991 e aggiornato nel 2002, raccolta quasi coincidente con quella statale.

Tali regole sono applicate e rispettate nella redazione dei testi normativi d'emanazione statale e regionale a partire dalla fine degli anni ottanta. Sfogliando i documenti normativi, non si può dire che tale applicazione sia stata, fino ad oggi, rigorosa ed uniforme da parte di tutti i

produttori di norme. Tuttavia, alcune analisi su testi a campione hanno verificato una crescente diffusione delle regole di tecnica legislativa.

B) Utilizzo di strumenti per il riconoscimento del linguaggio naturale

E' intuitivo che la presenza di regole condivise agevola la definizione di modelli testuali. E' altrettanto evidente che tale modellizzazione facilita il riconoscimento automatico di strutture dei testi normativi e la relativa marcatura secondo gli standard XML. Tale marcatura sarà infatti difficilmente ottenibile dal redattore dei testi normativi in quanto estranea ai compiti e alle finalità della sua attività. Se attuata da altri operatori con interventi successivi potrà provocare una dilatazione, spesso insostenibile, nei tempi e costi di costruzione e gestione di basi di conoscenza normative, strutturate secondo gli standard XML.

Proprio dalla prospettiva d'implementazione di un parsing efficiente per il riconoscimento automatico di strutture dei testi normativi e la successiva marcatura e conversione di tali testi in formato XML, prende avvio l'indagine che qui presentiamo.

2. Il metodo

Si può suddividere il percorso metodologico disegnato per tale indagine, nelle seguenti fasi:

1. individuazione dello strumento tecnico per l'implementazione del parser e l'estrazione delle informazioni;
2. individuazione e descrizione di modelli, in base a regole giuridiche;
3. individuazione e descrizione di strutture testuali che esprimono i modelli giuridici definiti;
4. scelta del campione di testi normativi da analizzare, compilazione della grammatica secondo la sintassi del parser prescelto e analisi automatica del campione.

Nei paragrafi successivi cercheremo di definire e descrivere tali fasi.

2.1. Individuazione dello strumento tecnico per l'implementazione del parser e l'estrazione delle informazioni

Lo strumento adeguato al riconoscimento e alla marcatura di strutture del testo normativo è stato individuato in Sophia 2.1 sistema di parsing che si avvale della metodologia ad automi a stati finiti e *finite state transducer*: software flessibile e configurabile che consente la formalizzazione di regole e modelli specifici (già definiti o in corso di definizione).

In particolare con il software suddetto si opera secondo le fasi seguenti:

- normalizzazione del testo in ingresso, marcando adeguatamente tutte quelle strutture e quei segmenti testuali riconoscibili sulla base dei caratteri, ovvero senza ricorso a consultazione del lessico-dizionario;
- analisi lessicale (attribuzione della categoria sintattica) e morfologica (tratti di flessione) del testo in ingresso;
- disambiguazione della categoria sintattica delle parole (*Part of Speech Tagging*);
- analisi sintattica parziale (denominata *chunking*), tesa ad identificare i gruppi sintattici minimi presenti nel testo in ingresso e a raggrupparli in costituenti;
- analisi semantica e identificazione delle strutture concettuali rilevanti nel testo d'ingresso;
- conversione del documento analizzato dal formato originario (Microsoft® Word, HTML, RTF, txt, ecc.) in formato XML, secondo la DTD stabilita.

2.2. Individuazione e descrizione di modelli in base a regole giuridiche

L'applicazione di un dispositivo di parsing come quello descritto, richiede la scrittura di un

set di regole per l'individuazione nei testi, di strutture linguistiche portatrici dell'informazione che si vuole estrarre. Si può parlare di compilazione di una grammatica specifica del dominio o del corpus di testi da analizzare.

La grammatica è composta da un insieme di modelli che definiscono le strutture linguistiche; a loro volta i modelli comprendono una o più regole che rappresentano una struttura linguistica e che vengono successivamente compilate secondo la sintassi del parser per l'analisi del testo e l'estrazione delle informazioni.

La linguistica computazionale indica come metodo più efficace per la costruzione della grammatica, l'estrazione dell'insieme di regole e modelli da un corpus di testi direttamente connessi con il dominio da trattare. In altre parole si cerca di ricostruire regole e modelli a posteriori, estraendo esempi di strutture linguistiche e classificandoli in base a valutazioni di tipo linguistico, logico, statistico, ecc..

L'applicazione aprioristica di tale metodica ai testi normativi, trascura però la specificità della natura e della funzione di tali testi.

Il testo normativo ha, per definizione, funzione prescrittiva, in altre parole incide sui comportamenti e lo status del destinatario, che non può sottrarsi a tale funzione.

In virtù di questa, si richiede (in base allo stesso principio di rappresentanza democratica, che legittima e al tempo stesso vincola colui che produce gli atti normativi, cioè il legislatore) che il testo normativo risponda ad un insieme di regole che sovrastano e contemporaneamente affiancano, integrano, a volte modificano le regole che compongono la lingua naturale (nel nostro caso l'italiano).

Tali regole si possono definire come regole giuridiche; la categoria è ampia e variegata; comprende al suo interno regole con una forza prescrittiva, che varia, in base alla fonte da cui provengono, a chi ne sono i destinatari, alle sanzioni che le corredano, ecc. Ad esempio, come vedremo, è la stessa Costituzione (che ha il rango più elevato tra le regole normative) a determinare che la delega legislativa debba contenere un termine entro il quale il Governo è tenuto ad emanare le norme oggetto di delega.

Si può stabilire una "graduatoria" della prescrittività di tali regole, partendo dalle fonti normative che, a loro volta, hanno diversi gradi di cogenza sulla base di quella che viene definita, appunto, gerarchia delle fonti.

Circa altre regole si discute se abbiano o meno valore normativo generale; tra questi vi sono le Circolari ministeriali o di altre autorità pubbliche (ad esempio dei presidenti del Parlamento); esse sono comunque senza dubbio regole giuridiche ed hanno forza prescrittiva nei confronti dei dipendenti di quel ministero o ente pubblico, essendo comunque atti amministrativi. Con tale tipologia di atto sono emanate, di solito, le regole di scrittura dei testi normativi, che qui assumono particolare rilevanza proprio per il fatto che hanno la specifica funzione di disciplinare la struttura del testo normativo.

Circa la validità e l'efficacia dei modelli costruiti in base a tali regole è opportuno aggiungere alcune considerazioni. Si possono elencare tre caratteristiche dalle quali dipende, in gran parte, la loro efficienza.

1. Flessibilità del modello, che deve adeguarsi alla molteplicità delle caratteristiche strutturali, funzionali e tematiche dei testi normativi, che utilizzano quel veicolo estremamente mutevole e imprevedibile che è il linguaggio e sono al servizio di un legislatore e di un interprete che necessitano di elasticità e discrezionalità per la formazione e l'applicazione della norma.
2. Precisione nella definizione del modello proprio grazie alla presenza di regole giuridiche in molti casi dettagliate e puntuali e per la prescrittività del testo normativo. Va da sé che conciliare flessibilità e precisione costituisce un punto

cruciale nella costruzione dei modelli stessi.

3. Autorevolezza della quale il modello, deve essere dotato, in modo che sia condiviso e ritenuto valido da tutti gli utenti del sistema. E' chiaro che tale autorevolezza è maggiormente garantita se il modello è derivato dalle regole normative propriamente dette, secondo la loro gerarchia mentre sarà garantita in misura minore da quelle regole che solo il redattore è tenuto a seguire; può essere più o meno garantita, in diverse misure, da modelli che si basano su prassi consolidate, teorie normative condivise da gran parte della dottrina, ecc. e via dicendo. Si può comunque dire che a modelli "più autorevoli" potrà corrispondere maggiore precisione, a modelli "meno autorevoli" dovrà corrispondere maggiore flessibilità.

Si può ancora osservare che modelli molto autorevoli hanno maggiore valenza comunicativa (si rivolgono e sono tendenzialmente accettati da tutti), mentre via via che l'autorevolezza diminuisce, i modelli assumono sempre più valenza interpretativa, sono cioè una definizione "soggettiva" della funzione di singole strutture testuali o dell'intero testo normativo.

Se si può quindi parlare di "Modelli comunicativi" e di "Modelli interpretativi" diventa fondamentale valutare la situazione d'impiego per scegliere se riferirsi ai primi o ai secondi.

Ad esempio, per la descrizione di un testo normativo, ai fini della comunicazione *erga omnes*, non si può che far riferimento ad un "modello comunicativo" e potrebbe risultare pericoloso e fuorviante l'uso di un "modello interpretativo".

2.3 Individuazione e descrizione di strutture testuali corrispondenti ai modelli giuridici definiti

Per l'implementazione della grammatica che poi verrà utilizzata dal parser è necessario integrare i modelli ricavati dalle regole giuridiche con le regole linguistiche. Bisogna precisare però che alcune regole linguistiche, in quanto disciplinano la scrittura del testo normativo, sono già recepite come regole giuridiche, in particolare come regole legislative e quindi la loro coerenza è rafforzata.

Si può dire ad esempio che le regole giuridiche prescrivono che un tipo della disposizione di modifica, si manifesti attraverso l'azione di sostituzione di parti di testo; la regola linguistica della sinonimia consente di dire che l'azione di sostituzione si esprime attraverso i verbi: sostituire, cambiare, mutare, ecc.. In questo caso la regola linguistica va ad integrarsi con le regole giuridiche, nella costruzione di un modello efficiente ai fini della funzione di parsing del testo. L'integrazione può tuttavia riguardare, anche casi in cui non tanto si integri la regola giuridica ma si completi la grammatica del testo normativo, andando a descrivere strutture linguistiche, cui non corrispondono regole giuridiche.

Dalla prassi testuale si possono ricavare regole che vanno a formare altri modelli di strutture linguistiche presenti nei testi normativi (o a integrare quei modelli ottenuti con le regole giuridiche). Modelli che sono estratti su base esemplificativa, ricorrendo all'analisi dei testi secondo la metodica praticata dalla linguistica computazionale.

E' chiaro che tali modelli non hanno la stessa precisione e soprattutto la stessa forza prescrittiva di quelli composti in base alle regole giuridiche. Li abbiamo quindi utilizzati e implementati nel parser solo come categoria residuale.

Chiamiamo questi ultimi modelli "malformati". Non tanto perché siano scorretti da un punto di vista linguistico, ma per contrapporli a quelli che corrispondono alle regole giuridiche, indicate nel paragrafo precedente, che definiamo "benformati".

I testi normativi contengono comunque strutture linguistiche che non corrispondono ai modelli descritti, perché i testi stessi possono contenere veri e propri errori giuridici e

linguistici.

In questi casi completiamo la grammatica da implementare nel parser, con regole (le definiamo “casistiche”) che rappresentano eccezioni ai modelli: si tratta o di veri e propri errori o di eccezioni alla regola, come lo sono le eccezioni nelle grammatiche linguistiche.

2.4. Compilazione della grammatica e analisi automatica del campione

La compilazione della grammatica nella sintassi propria del parser scelto, avviene utilizzando il Workbench del sistema descritto nel paragrafo 2.1., attraverso la scrittura di regole di produzione, che formalizzano i modelli definiti e consentono l'individuazione automatica delle strutture linguistiche descritte e l'estrazione delle informazioni.

La scelta del campione di testi normativi da analizzare deve ovviamente rispondere a criteri rappresentativi del dominio linguistico normativo, dal quale s'intende estrarre le informazioni. Descrivendo l'indagine svolta, indicheremo i criteri seguiti per l'individuazione del campione oggetto di questa prima analisi.

L'analisi del campione può essere poi compiuta in fasi successive in modo da:

- valutare i risultati ottenuti;
- integrare e modificare i modelli benformati definiti a priori;
- individuare e formalizzare modelli malformati e regole casistiche;
- estendere l'analisi su un *corpus* via via più ampio per verificare l'efficienza del sistema di parsing.

3. Prime analisi e applicazioni

Abbiamo deciso di sperimentare il metodo descritto per il riconoscimento automatico e l'estrazione di tre strutture tipiche dei testi normativi, strutture che rappresentano:

- la disposizione di delega legislativa;
- la disposizione di modifica testuale esplicita o novella;
- il rinvio esplicito testuale esterno.

Tra le molteplici strutture che compongono un testo normativo e che sono definite da altre regole giuridiche, la scelta è caduta sulle strutture elencate sopra, principalmente per due ordini di considerazioni.

1. Si tratta di strutture puntualmente definite, sia giuridicamente, sia linguisticamente dalle regole legislative e da altre regole normative molto vincolanti.
2. Svolgono la funzione principale di nessi normativi, cioè di collegamento tra testi normativi diversi, funzione particolarmente rilevante sia ai fini di qualsiasi intervento o utilizzo dei testi (si pensi alla compilazione del testo coordinato), sia ai fini della ricostruzione della dimensione intertestuale, che, insieme a quella contestuale è fondamentale in un'analisi di tipo linguistico-testuale. E' nostra intenzione, infatti, proseguire l'indagine, tentando l'analisi e la formalizzazione di altre strutture normative che esprimono nessi, come la proroga o la sospensione.

Nei paragrafi seguenti presentiamo i primi risultati raggiunti che riguardano:

- per la delega legislativa l'analisi giuridica per la costruzione del modello;
- per la novella la modellizzazione della struttura testuale, in base a regole normative e linguistiche;
- per il rinvio testuale esplicito l'applicazione di tutte le fasi metodologiche descritte.

4. La delega legislativa

La delega legislativa dal Parlamento al Governo è prevista dalla Costituzione (articolo 76), che detta le regole, alle quali la stessa disposizione di delega deve uniformarsi:

- deve esser fatta al Governo;
- deve contenere un termine entro cui il Governo deve emanare l'atto delegato;
- deve specificare l'oggetto della delega;
- deve contenere i principi e i criteri direttivi ai quali il Governo deve attenersi nell'esercizio della delega medesima.

Il Parlamento può dettare condizioni che il Governo deve rispettare (richiesta di pareri, di audizioni, emanazione di altri atti ecc.). Tali condizioni non sono indicate esplicitamente dalla Costituzione, ma dottrina e giurisprudenza ritengono che il Parlamento sia libero d'inserirle nella delega.

E' chiaro che tali regole, essendo dettate dalla Costituzione hanno il massimo di forza cogente. Un testo normativo che non le rispettasse sarebbe impugnabile di fronte alla Corte costituzionale e possono essere cambiate solo con una modifica alla Costituzione e non da una legge ordinaria o da altri atti normativi.

Una legge ordinaria (la legge 23 agosto 1988, n. 400) specifica poi le regole costituzionali e aggiunge modalità che alcune particolari tipologie di deleghe devono rispettare. In particolare l'articolo 14, commi 3 e 4 recita:

"3. Se la delega legislativa si riferisce ad una pluralità di oggetti distinti suscettibili di separata disciplina, il Governo può esercitarla mediante più atti successivi per uno o più degli oggetti predetti. In relazione al termine finale stabilito dalla legge di delegazione, il Governo informa periodicamente le Camere sui criteri che segue nell'organizzazione dell'esercizio della delega.

4. In ogni caso, qualora il termine previsto per l'esercizio della delega ecceda i due anni, il Governo è tenuto a richiedere il parere delle Camere sugli schemi dei decreti delegati. Il parere è espresso dalle Commissioni permanenti delle due Camere competenti per materia entro sessanta giorni, indicando specificamente le eventuali disposizioni non ritenute corrispondenti alle direttive della legge di delegazione. Il Governo, nei trenta giorni successivi, esaminato il parere, ritrasmette, con le sue osservazioni e con eventuali modificazioni, i testi alle Commissioni per il parere definitivo che deve essere espresso entro trenta giorni."

Tale disposizione aggiunge almeno tre regole alla disciplina della struttura della disposizione di delega:

- la delega può contenere più oggetti distinti;
- il Governo può emanare più atti delegati in caso di più oggetti;
- se il termine previsto per l'esercizio della delega è maggiore di due anni, il Governo deve richiedere il parere delle Camere secondo una procedura che è definita.

Inoltre dal comma 1 dello stesso articolo 14 si ricava che l'atto (o gli atti) con cui il Governo esercita la delega deve denominarsi "decreto legislativo".

La forza cogente di tali regole non è più quella della Costituzione, ma di una legge ordinaria. Dunque può essere modificata da un'altra legge ordinaria. Un giudice potrebbe disapplicare, per il caso sottoposto al suo sindacato, una disposizione di delega ritenendola in contrasto con quanto dettato dalla legge 400/1988, ma la Corte costituzionale non potrebbe dichiarare illegittima (e quindi nulla per tutti), una disposizione di delega non rispondente ai dettami della legge 400/1988.

Si può ancora ricordare che la regola utilizzata è del 1988: non vale quindi per le disposizioni di delega precedenti, come non valgono le regole costituzionali per le deleghe

precedenti al 1948. La definizione delle strutture di delega anteriori a tali date sarà quindi meno cogente o dovrà rifarsi ad altre regole.

Come detto le regole di redazione dei testi normativi per Parlamento e Governo sono contenute nella Circolare del 20 aprile 2001, n. 10888. In particolare l'articolo 2 si occupa anche della delega legislativa e disciplina puntualmente la struttura di tale disposizione:

"Le disposizioni contenenti deleghe legislative, ai sensi dell'articolo 76 della Costituzione recano i seguenti elementi: 1) il destinatario della delega (il Governo); 2) il termine per l'esercizio della delega e l'eventuale termine per l'emanazione di disposizioni integrative o correttive; 3) l'oggetto della delega; 4) i principi e i criteri direttivi (che devono essere distinti dall'oggetto della delega). Il termine "delega" è usato esclusivamente in presenza di una delegazione legislativa con la formula: "Il Governo è delegato ad adottare...". E' inoltre sempre indicata la denominazione propria dell'atto da emanare (decreto legislativo) ed è precisato se la delega può essere esercitata con uno o più atti. Le disposizioni di delega sono contenute in un apposito articolo. Un articolo non può contenere più di una disposizione di delega".

Abbiamo già accennato al valore normativo e alla forza prescrittiva di una circolare, di grado senz'altro inferiore a quelle della Costituzione e della legge ordinaria.

E' comunque altrettanto evidente che le leggi le scrivono (o le correggono) i drafter, che, per dovere del loro ufficio, conoscono e rispettano le regole di tecnica legislativa.

Vale la pena riproporre l'osservazione, di carattere temporale, già fatta a proposito della Costituzione e della legge 400/1988; la Circolare è addirittura del 2001 e quindi i vincoli giuridici e linguistici da essa imposti alla struttura della delega, troveranno riscontro solo nei testi legislativi più recenti e, via via che si va indietro nel tempo, ci si può imbattere in strutture variabili.

Tenuto conto dei limiti prescrittivi e temporali, cui va soggetto l'insieme di regole che abbiamo provato a individuare e descrivere, si può sostenere che esse formano un modello benformato da un punto di vista giuridico della struttura della delega legislativa.

Sulla base di tale modello si può provare a descrivere il testo inserendo dei marcatori o qualificatori degli elementi che compongono la struttura come nel seguente esempio:

<DESTINATARIO> Il Governo della Repubblica <DESTINATARIO> <AZIONE DI DELEGA> è delegato ad emanare, <AZIONE DI DELEGA> <TERMINE> entro diciotto mesi dalla data di entrata in vigore della presente legge <TERMINE>, <ATTO/I DELEGATO/I> uno o più decreti legislativi <ATTO/I DELEGATO/I> <OGGETTO DELEGA> recanti disposizioni integrative della legislazione in materia di tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali, <OGGETTO DELEGA> <CRITERI DIRETTIVI> con l'osservanza dei seguenti principi e criteri direttivi: a) specificare le modalità di trattamento dei dati personali utilizzati a fini storici, di ricerca e di statistica, tenendo conto dei principi contenuti... <CRITERI DIRETTIVI> (Legge 31 dicembre 1996, n. 676).

Una tale descrizione dovrebbe successivamente consentire di specificare e dettagliare, anche da un punto di vista linguistico, le strutture individuate sulla base delle regole giuridiche e di passare poi alla compilazione delle regole per l'implementazione del parser e all'analisi del corpus, come vedremo nei paragrafi successivi, per le altre strutture oggetto dell'indagine.

5. La modifica testuale esplicita o novella

Le disposizioni di modifica, secondo Sartor, rientrano fra le principali tipologie di nessi normativi, classificate in base al loro impatto sulle norme interessate. Le modifiche distinte dall'altro grande ramo dei rinvii o riferimenti, sono nessi normativi caratterizzati dal fatto che la

disposizione attiva incide sulla disposizione passiva, eliminandola, cambiandone il testo o cambiandone la portata normativa (pur lasciandone invariato il testo). Questa incidenza manca invece nel rinvio, in cui la disposizione attiva si avvale della disposizione passiva, per completare il proprio significato, senza influire su quest'ultima.

In relazione alla natura dell'impatto della modifica sulla disposizione passiva, distinguiamo tra modifiche testuali, modifiche temporali (incidono sull'ambito temporale di applicabilità della disposizione passiva), modifiche materiali, (modificano il contenuto normativo della disposizione passiva senza incidere sul testo). Noi ci occupiamo solo delle prime, le modifiche esplicite testuali che, tradizionalmente, i giuristi chiamano *novelle*.

Anzi è forse più corretto dire che la funzione della modifica normativa esplicita si esprime attraverso tre aspetti:

la struttura della novella, che si compone di una parte introduttiva, detta *alinea* e una parte che contiene la *modifica testuale esplicita*.

Le caratteristiche dell'atto normativo modificante e dell'atto modificato: indispensabili per poter successivamente ricostruire i nessi di modifica fra le diverse fonti normative.

La citazione con cui si richiama il documento da modificare, che esprime il riferimento normativo (testuale esplicito anch'esso), elemento portante delle disposizioni di modifica.

In base ai tre aspetti indicati abbiamo cercato di definire e descrivere gli elementi qualificanti la disposizione di modifica. Tale descrizione, che riportiamo qui di seguito, è ricavata dalle regole di tecnica legislativa e dall'analisi di un campione di circa 100 disposizioni di modifica contenute in 8 provvedimenti normativi statali (le quattro leggi cd. Bassanini, e altri testi ad esse collegati), emanati tra il 1968 e il 1999.

Tipo di atto modificante: indica la tipologia (legge, decreto-legge, d.p.r., decreto legislativo, ecc.-) dell'atto normativo in cui è contenuta la modifica; serve a ricostruire velocemente i nessi fra le norme in presenza di modifiche ed è composta da:

Nome dell'atto, Data, Numero: indicano gli estremi dell'atto normativo modificante, sia nel formato di citazione esteso, sia nel formato semplificato.

Posizione novella: è la posizione, all'interno del testo modificante, in cui si trova la disposizione di modifica, in modo da identificare la formula di modifica con precisione; e anche di evidenziare subito a quale livello della struttura essa si presenti.

Oggetto della modifica: indica l'oggetto della modifica in senso stretto (ovvero se la modifica vada ad incidere sull'intera legge, oppure su una partizione, quale articolo, comma, ecc.). Tale elemento è importante anche dal punto di vista della struttura, perché quando si va a modificare una partizione, l'effetto della modifica si riflette anche su quella di livello immediatamente superiore, in particolare per le aggiunte o le abrogazioni.

Tipo di atto da modificare, composto da:

Nome, Data, Numero dell'atto da modificare : con questi elementi si indicano le caratteristiche dell'atto normativo che si va a modificare, il tipo di atto e gli estremi del documento.

Azione: l'elemento descrive l'azione di modifica; dovrebbe assumere solo i valori standard, a volte combinati fra loro: *abrogazione, sostituzione, inserimento, aggiunta*, ma può assumere valori diversi (ad. es.: *ricollocamento*).

Espressione: è la forma linguistica con cui viene disposta la modifica, racchiusa fra le virgolette o altri eventuali segni ortografici (due punti, parentesi, ecc.), che delimitano le modifiche. L'espressione contiene l'enunciato che dispone l'azione, fino ai due punti che introducono il nuovo testo.

Il testo della modifica: che, in base alle regole di redazione, è racchiuso tra virgolette e preceduto da due punti.

Si sono inoltre individuati e definiti alcuni elementi testuali (preposizioni, avverbi, congiunzioni, ecc.), che fungono da connettori e qualificatori dei vari elementi della disposizione di modifica.

6. Il rinvio esterno testuale esplicito

6.1. La citazione normativa

La citazione esprime il riferimento o rinvio normativo esplicito, elemento costitutivo della disposizione di modifica, ma anche elemento indipendente con funzione di nesso che integra disposizioni diverse, pur non modificandone il testo. Una sua rigorosa formalizzazione è dunque indispensabile sia per il riconoscimento della modifica esplicita, sia per ogni attività di coordinamento automatico tra testi normativi.

Abbiamo individuato quali elementi costitutivi e distintivi della citazione normativa:

1. la **Partizione** che indica qualsiasi parte esplicitamente contrassegnata da una particolare espressione grafica, in cui si suddivide l'atto (ad es. articolo, comma, lettera, capo, titolo, ecc.) con il relativo segno grafico di enumerazione (es. 3, c), ottavo, A, ecc.);
2. l'**Atto (nome+data+numero)**. Per Atto s'intende l'identificatore univoco dell'atto normativo richiamato, che di regola, si esprime con il nome ufficiale dell'atto (es. legge, decreto-legge, decreto ministeriale, ecc.), la data di promulgazione o emanazione e il numero d'ordine assegnato all'atto.

Circa il nome sono opportune due osservazioni:

- il nome, per essere univoco, necessita spesso di alcune specificazioni da aggiungere a quello che è il nome proprio dell'atto; ad esempio per citare un decreto ministeriale, si dovrà indicare il dicastero del ministro che l'ha emanato, pena una non univoca individuazione dell'atto citato;

- di solito quando si cita un atto normativo, contenuto in un altro atto, si cita con il nome del contenitore. Ad esempio un regolamento non si cita come tale, ma come Decreto del Presidente della Repubblica o Decreto del Ministro. Non sempre, però, questa regola è applicata (vedi, ad esempio, la citazione dei codici o dei testi unici).

Inoltre la citazione normativa può essere integrata con alcuni elementi che connettono le parti della citazione (in particolare preposizioni e punteggiatura). E' prassi frequente, ad esempio, usare la preposizione per connettere la citazione delle partizioni con quella dell'atto (ad es.: l'articolo 20 *della* legge...).

In base alla diversa scrittura dell'elemento Atto, abbiamo classificato i vari formati di citazione previsti dalle regole di redazione, in tre categorie, denominate: *normale*, *semplificata*, *non paradigmatica*.

6.2 Modello di citazione normale

Nello scrivere le regole che hanno implementato il modello del riferimento sia nella sua accezione benformata che in quella malformata, abbiamo usato soprattutto tre moduli del Workbench di Sophia 2.1: il tokenizer, il compounder ed il semantics.

Nel primo si formalizzano regole per la descrizione di stringhe di caratteri (es.: n. 25 del 2001 = (n|N)". "" "[0-9][0-9]*" "(del|DEL)" "[0-9][0-9][0-9][0-9]).

La grammatica del parser a questo livello (ma anche nel livello semantics) usa vari operatori tra i quali i più ricorrenti sono:

?	Facoltatività
*	0 o infinite volte

Nel modulo compounder abbiamo definito i sintagmi nominali (decreto legge, legge costituzionale, codice penale, codice civile, ecc.).

Ma è nel modulo semantics che è possibile scrivere le regole necessarie per l'analisi semantica e l'identificazione delle strutture concettuali rilevanti nel testo di input.

In tale modulo infatti è possibile costruire le regole che compongono i modelli da estrarre. Il pattern di tale modulo sarà costituito da categorie semantiche precedentemente assegnate. Tali categorie semantiche interagiranno anche con categorie lessicali, morfologiche e sintattiche, così come con macro in precedenza definite all'interno del modulo semantico stesso.

Naturalmente ad ogni valore attribuito alla categoria semantica dovrà corrispondere una variabile, indispensabile per l'estrazione del dato rilevante e quindi per la creazione del template.

6.2.1 L'Atto.

Innanzitutto la citazione deve essere univoca per individuare quello, e solo quello, specifico atto a cui si riferisce. Quasi sempre questa univocità si ottiene riportando tre informazioni:

il *nomen* dell'atto (che indica la categoria a cui l'atto appartiene).

Per individuare tutti i possibili nomi di atto abbiamo costruito una macro [M-NOME-ATTO] la cui formalizzazione è stata definita nel modo seguente:

NOME-ATTO:vnome+((DI+DETX)?+AUTORITA:vlautorita+DI?+DETX?+[M-QUALIFICA])?)+DI?+DETX?.

La macro è composta da tre elementi:

il nome dell'atto vero e proprio;

il nome dell'autorità emanante, quando il nome dell'atto individua esplicitamente tale elemento (ad es. in decreto legislativo è implicito che l'autorità emanante è il Presidente della Repubblica, mentre per il Decreto del Presidente della Repubblica è esplicitata l'autorità emanante);

la categoria semantica QUALIFICA che serve ad individuare la singola autorità emanante, quando questa è connotata da qualifiche di luogo (es.: legge regionale Toscana) o di materia (es.: decreto del Ministro dell'ambiente).

La data, di solito, quella di promulgazione o di emanazione (il formato standard è: giorno (in cifre)/mese (in lettere)/anno (in cifre)).

Il numero che identifica l'atto normativo (ha un massimo di quattro cifre) ed è stato formalizzato usando due categorie semantiche: PARTNUM alla quale è associata l'abbreviazione della parola numero ((n|N)".") e la parola "numero" stessa e NUMBER, categoria alla quale sono associate regole che prevedono sia i numeri arabi che quelli romani.

A titolo d'esempio vediamo come si esprime il modello della citazione normale nel modulo di formalizzazione delle struttura semantica, nel quale si utilizzano elementi definiti dalle regole di produzione dei livelli inferiori (lessicale, morfologico e sintattico).

New-atto	[M-NOME-ATTO]+ (AN-NUM-ATTO:vabbrevdata (DATE:vdata+(PUNCTX)?+PARTNUM+(NUMBER:vnum)))+[M-ALLCAT-END]	{actionNEW(vnome:Riferimento:"Cit_Atto", vnome:attonome:head, vlautorita:attoautorita:alllist, vlqualifica:attoqualifica:alllist, vdata:attodata:all, vabbrevdata:attoabbrevdata:all, vnum:attonumero:head, vnome:rifBeg:start,
----------	--	---

Tale formalizzazione consente di identificare le citazioni normali, di generare in output un template identificativo del nome dell'atto, dell'autorità emanante, della sua eventuale qualifica, della data e del numero dell'atto. Una volta identificati tali valori è possibile inserire i dati estratti in un database oppure è possibile la marcatura XML. A tal proposito si noti la macro finale [M-ALLCAT-END] provvisoriamente inserita come tag XML di chiusura.

6.2.2 Testi unici

Un caso particolare di citazione normale riguarda i testi unici. Il testo unico è citato, secondo le regole, facendo seguire il nome “*testo unico*” dal titolo e dalla citazione dell'atto normativo che lo contiene e la sua formalizzazione è avvenuta nel modo seguente:

(TESTO-UNICO:vstart)+(DI+DETX)?+(DISP-TRANS:vnome|NOME-ATTO:vnome)?
+DI?+[M_ALLCAT]+((DI+PX+DETX)|(FONTE-COGNIZIONE+PX))+[M-ATTO]+[M-ALLCAT-END]

Una forma corretta di citazione di testo unico può essere la seguente: “*testo unico delle disposizioni per l'assicurazione obbligatoria contro gli infortuni sul lavoro e le malattie professionali, approvato con decreto del Presidente della Repubblica 30 giugno 1965, n. 1124.*”

La categoria semantica TESTO-UNICO tesa ad identificare il nome dell'atto, può essere seguita dalle categorie semantiche DISP-TRANS e NOME-ATTO, inserite per prevedere i casi in cui la citazione contenga le parole “legge” o “disposizione” alle quali, per la formalizzazione di altre regole, sono state attribuite specifiche categorie semantiche. Per tutte le altre parole, che possono trovarsi tra il nome del testo unico e l'atto contenitore, è stata usata la macro M_ALLCAT, mentre il pattern ((DI+PX+DETX)|(FONTE-COGNIZIONE+PX)) ha la funzione di riconoscere frasi come “*di cui al*” o “*approvato con*”.

6.2.3 Atti senza numero

Un altro caso particolare di citazione normale riguarda gli atti non numerati. Secondo le regole regionali di tecnica legislativa, per la citazione di tali atti si devono aggiungere gli estremi di pubblicazione in G.U. o sul bollettino regionale, oppure se ne ricorda il titolo o altri elementi idonei ad identificarli. La regola d'estrazione ha tenuto conto solo della formalizzazione degli estremi di pubblicazione, ma avrebbe potuto contenere la macro M_ALLCAT per descrivere il titolo dell'atto o altri elementi tesi all'identificazione dello stesso:

[M-NOME-ATTO]+DATE:vdata+(PUNCTX)?+[M-PUBBLICAZIONE]?+[M-ALLCAT-END].

La macro PUBBLICAZIONE l'abbiamo inserita come opzionale, in ossequio al criterio di formalizzare regole, sì rispettose delle norme di tecnica legislativa, ma con una qualche flessibilità che trova giustificazione, come accennato sopra, nei testi analizzati.

6.3 La partizione

Prima di illustrare la formalizzazione dei modelli delle citazioni semplificate e non paradigmatiche, vediamo l'altro elemento costitutivo della citazione normativa: la partizione.

6.3.1 Articoli e sottopartizioni

Il riferimento ad una partizione dell'atto può completare la citazione.

Per partizione si intende qualsiasi parte, esplicitamente contrassegnata da una particolare

espressione, in cui si suddivide l'atto (numero, lettera, comma, articolo, sezione, capo, ecc.).
Le regole stabiliscono che la citazione di più partizioni in ordine gerarchico sia in ordine discendente (articolo, comma, lettera...).

New-partizione	(DETX)?+ARTICOLO:vstart+[M-COORD-1ART]+ (PUNCTX+DETX?+((([M-ORDINALE-COOR]+COMMA+PUNCTX?) (COMMA+[M-COORD-1COM])))?) +(PUNCTX+ALINEA:valinea)?+ (PUNCTX+DETX?+LETTERA+[M-COORD-1LET])?+(PUNCTX+ALINEA:valinea)?+ (PUNCTX+DETX?+PARTNUM+[M-COORD-1NUM])?+ PUNCTX?+DI?+([M-ATTO] [M-TESTO-UNICO])+ [M-ALLCAT-END]	{actionNEW(vstart:Riferimento:"Partizione_simpl e", vspec:specificazione:all, vlart1:1articoli:alllist, vlra1:1r_art:alllist, vlcom1:1commi:alllist, vlrc1:1r_com:alllist, vlet1:1lettere:alllist, vlr11:1r_let:alllist, vlnum1:1partnumeri:alllist, vlrn1:1r_partnum:alllist, vnome:attonome:head, vsigla:attonomesigla:all, vlautorita:attoautorita:alllist, vlqualifica:attoqualifica:alllist, vdata:attodata:all, vabbrevdata:attoabbrevdata:all, vnum:attonumero:head, vcodici:attocodici:all, valinea:alinea:all, vstart:rifBeg:start, vend:rifEnd:end)}
----------------	---	---

La regola è formata dalla descrizione della partizione *Articolo* (prevista come necessaria) e delle relative sottopartizioni (*comma*, *lettera*, *numero*) alle quali si sono aggiunte due macro ([M-ATTO]||[M-TESTO-UNICO]) che rappresentano tutte le possibili forme di citazione di atti.

In tal modo si ottiene la formalizzazione dell'intera citazione, sia nella sua forma normale che in quella semplificata.

La categoria semantica ARTICOLO comprende sia la parola "articolo" sia l'abbreviazione "A|art." mentre la macro [M-COORD-1ART] identifica sia un solo numero, sia più numeri.

La citazione delle partizioni inferiori all'articolo segue lo stesso schema di formalizzazione. Si nota però:

3. la citazione del comma comprende anche la macro [M-ORDINALE-COOR] che formalizza i casi in cui si citino uno o più commi non numerati;
4. abbiamo inserito anche l'alinea che può essere citata come elemento a sé stante della partizione;
5. non sono, per ora, previste altre partizioni come il *paragrafo* usato nel codice civile, o le partizioni inferiori al numero, che però possono essere facilmente formalizzate in base allo schema descritto.

Il riferimento può riguardare una o più partizioni del medesimo atto. Nel secondo caso possiamo parlare di citazione multipla.

La regola appena illustrata estrae citazioni multiple del tipo:

articoli 5, 6 e 7 della legge 25 giugno 1988, n. 315;
articoli dal 3 al 12 della legge 25 giugno 1988, n. 315;
articolo 20, commi 2 e 3 della legge 25 giugno 1988, n. 315,

ma non comprende altri tipi di citazioni multiple, dove viene ripetuto il nome della partizione due o tre volte.

Usiamo l'espressione partizione multipla, perché, di fatto, questa forma di scrittura contiene non uno, ma due o più riferimenti a partizioni diverse del medesimo atto. Al fine di consentire operazioni per il collegamento e il coordinamento automatico dei testi normativi, sembra opportuno considerare e qualificare la citazione multipla come più riferimenti distinti, secondo lo schema:

partizione1+nome atto+data+numero.....partizioneN+nome atto+data+numero

Tutte le macro relative alle partizioni sono state definite, tenendo conto di questo fenomeno della citazione multipla; mentre l'output di Sophia, a livello di template, distingue le singole partizioni dell'atto, ciò non è stato ancora previsto nella marcatura del file XML relativo, nel quale si individua solo il primo elemento della partizione e l'ultimo elemento dell'atto citato. Ciò non esclude che questo lavoro di marcatura XML delle singole partizioni non possa essere fatto con semplicità in seguito.

Per formalizzare tali casi non abbiamo fatto altro che creare altre due regole in cui lo schema ARTICOLO COMMA LETTERA PARTNUM ([M-ATTO]|[M-TESTO-UNICO]) viene ripetuto 2 o 3 volte (chiaramente ad ogni elemento sono associate le relative variabili).

Sarebbe preferibile potere formalizzare in una sola regola il modello delle partizioni multiple anziché scomporlo in tante sotto regole.

Ciò per almeno due motivi:

1. in questo come in altri casi la "formalizzazione su base casistica" anziché su modello generale, tende a limitare l'efficienza della regola per l'estrazione dell'informazione (in questo caso per esempio non si individuano le partizioni multiple col nome della sottopartizione ripetuto più di tre volte, se non aggiungendo un'ulteriore regola o ulteriori elementi a una regola);
2. considerazioni di ordine metodologico e di economia generale imporrebbero che il sistema fosse efficiente col minor numero di regole possibili e con regole molto semplici.

6.3.2 Sovrapartizioni

Pur essendo possibile formalizzare in un'unica regola tutti i riferimenti con sovrappartizioni, abbiamo dovuto scrivere distinti ed autonomi *patterns* per il riconoscimento di queste stringhe, al fine di consentire l'individuazione dell'inizio della citazione per la marcatura XML. Il tag di inizio della marcatura infatti non tollera elementi in or tra loro (Libro|Parte|Titolo|Capo).

6.3.3 Annessi

Talora una parte del contenuto dell'atto normativo, non è strutturata in articoli o comunque si presenta in un allegato. Questo si può manifestare in diverse forme: da semplici elementi testuali, a complesse tabelle, ad interi documenti strutturati. Le tabelle, i quadri, i prospetti e gli elenchi non vanno inseriti nel testo degli articoli né in allegato ad essi, ma in allegato al testo normativo, dopo l'ultimo articolo. All'inizio di ciascun allegato deve essere citato l'articolo, o il primo articolo, che rinvia all'allegato stesso. Tale citazione non è richiesta nel caso di allegato contenente le modificazioni apportate in sede di conversione a decreti-legge. Gli allegati non contengono nelle note esplicative, ulteriori disposizioni sostanziali. Queste ultime trovano collocazione esclusivamente nell'articolato.

Mentre nel Manuale regionale di tecnica legislativa, la struttura formale dell'allegato è puntualmente definita e di conseguenza anche la forma di citazione risulta ben determinata,

invece la circolare della Camera dei Deputati del maggio 2001, non definisce con altrettanto rigore la struttura dell'annesso.

Ciò determina una necessaria flessibilità nella formalizzazione della citazione degli allegati che abbiamo rappresentato con la regola (*New-allegato*) che segue.

```
(ALLEGATO:vlann)+[M-NUM-COOR]?+[M_ALLCAT]+  
(PUNCTX+ALLEGATO:vlann+[M-NUM-COOR])*+  
PUNCTX?+(AX|ALLEGATO)?+  
(DI|PX|PUNCTX)?+([M-ATTO]||[M-TESTO-UNICO])+  
[M-ALLCAT-END]
```

La regola si basa appunto, in primo luogo sulle indicazioni del Manuale e anche sulla prassi diffusa nei testi statali che danno vita alla seguente struttura:

ALLEGATO - PARTIZIONE ALLEGATO - CONNETTORE - NOME DELL'ATTO NORMATIVO.

La suddetta flessibilità nella forma di citazione dell'allegato, si è tradotta in una regola che prevede molti elementi in forma opzionale e addirittura la macro ALLCAT per individuare diversi termini.

6.4 La citazione multipla

Possiamo distinguere tra citazioni multiple di partizioni e citazioni multiple di atti.

Nel primo caso si tratterà della citazione di più partizioni del medesimo atto; la formalizzazione di tale categoria è stata già illustrata nel paragrafo 6.3.1.

Nel secondo caso invece si intende descrivere i casi in cui con una sola citazione si fa riferimento a più atti.

Le regole di tecnica legislativa non menzionano questa particolare forma di citazione, che noi tuttavia riteniamo regolare, sia perché non espressamente vietata dalle regole stesse, sia perché corretta dal punto di vista delle regole linguistiche.

Non disponendo di un modello predefinito è stata dunque l'analisi del nostro campione a suggerirci due forme distinte di citazione multipla di atti.

Una prima forma può essere: “*regi decreti 16 marzo 1942, numeri 699 e 700*” secondo lo schema ATTO + DATA + (NUMERO)*.

Il secondo caso prevede invece la seguente forma di citazione: “*decreti del Presidente della Repubblica 14 gennaio 1972, n. 1, n. 2, n. 3, n. 4, n. 5, n. 6, 15 gennaio 1972, n. 7, n. 8, n. 9, n. 10, n. 11 e 5 giugno 1972, n. 315*”. Tale citazione è riconducibile al modello ATTO + (DATA + NUMERO)*.

6.5 Modello di citazione semplificata

Le regole di tecnica legislativa prevedono l'uso di citazioni abbreviate o semplificate nel corpo del testo dell'atto normativo.

Sono previsti due tipi di citazione *semplificata* che si possono formalizzare nel modo seguente:

6. Nome atto (esteso o abbreviato) + Numero atto/Anno emanazione-promulgazione (es.: l. 400/1988 o legge 400/1988);
7. Nome atto (esteso) + **n.** + Numero atto + **del** + Anno emanazione-promulgazione (es.: legge n. 400 del 1988).

Pressochè tutte le sigle corrispondenti al nome di un atto abbreviato, sono state definite nel tokenizer e classificate nella categoria semantica ATTO-SIGLA che è stata poi inserita nella regola *new-atto-abbrev* che segue:

pattern	action
---------	--------

ATTO-SIGLA:vsigla+ ([M-QUALIFICA])?+ AN- NUM-ATTO:vabbrevdata+ [M-ALLCAT-END]	{actionNEW(vend:Riferimento:"Atto_abbrev", vsigla:attonomesigla:all, vlqualifica:attoqualifica:allist, vabbrevdata:attoabbrevdata:all, vsigla:rifBeg:start, vend:rifEnd:end)}
--	--

Per la citazione di due casi, in cui i nomi atto abbreviati (del. e d.) vengono evidenziati nel template rispetto alle autorità emananti (d. pres. giunta reg. 1/1989) e per i casi in cui non è semplificato il nome dell'atto ma la data e il numero, il sistema si serve della regola *new-atto*.

6.6 Modello di citazione non paradigmatica

La citazione di alcuni atti normativi, tra i quali i codici e la Costituzione, non segue il modello normale (nome atto+data+numero), ma per prassi consolidata accolta dalle regole di redazione, è espressa con l'indicazione della partizione, seguita dal solo nome esteso dell'atto. In tali casi lo schema è dunque il seguente: partizione+preposizione (facoltativa) + nome atto. Es.: *articolo 345 del codice civile; articolo 3 Costituzione*.

Abbiamo formalizzato questo tipo di citazione con la presenza necessaria della partizione davanti al nome dell'atto, ritenendo che la citazione di atti come la Costituzione o il codice civile, senza il riferimento esplicito ad un articolo possa non essere o non sia generalmente un rinvio o un nesso (che comunque così espresso è del tutto generico), ma possa avere altri significati e funzioni all'interno del testo normativo.

La categoria semantica CODICI è stata dunque inserita all'interno della macro M-ATTO, che come abbiamo visto sopra, opera in tutte le regole, che prevedono una partizione o sovrartizione.

Abbiamo tuttavia anche previsto la regola *new-codici*

[M-DISP-TRANS]+PUNCTX?+DI?+DETX?+CODICI:vcodici+[M-ALLCAT-END]

con la quale si è voluto formalizzare la citazione del tutto peculiare delle *disposizioni transitorie* della Costituzione.

Si ritiene che tale regola possa essere utilizzata anche per l'estrazione della citazione delle disposizioni preliminari al codice civile (chiamate anche *disposizioni sulla legge in generale*, per ora inserite nella categoria semantica CODICI) e di altre forme di citazione assimilabili a tale schema.

6.7 Citazioni malformate

I modelli (e sottomodelli) appena descritti, consentono di riconoscere le citazioni che possiamo chiamare benformate secondo le regole della tecnica legislativa.

Come sappiamo (cfr. par. 1) nei testi normativi le regole formali di scrittura della citazione non sono costantemente rispettate. Per questo motivo, sulla base dei risultati dell'analisi del corpus normativo selezionato, abbiamo compilato regole per riconoscere quelli che possiamo definire, formati irregolari di citazione o *citazioni malformate*.

E' però intuitivo che la formalizzazione delle citazioni irregolari, indeterminabili per definizione, presenta notevoli difficoltà. In particolare si può considerare assai difficoltoso il riconoscimento di quelle citazioni (che possiamo chiamare scorrette) che non soddisfano nemmeno le condizioni minime di univocità del riferimento normativo, facendone venire meno persino la funzione.

Rispetto alle citazioni malformate ci siamo comportati nel modo seguente.

In base all'analisi del campione abbiamo individuato a posteriori, citazioni irregolari ricorrenti, con scostamenti più o meno ampi rispetto al modello normale. Citazioni che abbiamo poi classificato nelle seguenti categorie:

citazione invertita e citazione con partizione invertita;
 citazione con abbreviazioni e scioglimenti scorretti;
 citazione con elementi mancanti o pleonastici;
 citazione con errori ortografici;
 citazioni *storiche*.

Non si esclude che analizzando un corpus sempre più ampio, sia possibile individuare altri casi di citazioni malformate, non riconducibili a tali categorie.

Inoltre, per il momento, non è stato possibile tradurre ciascuna delle categorie individuate in un corrispondente modello formalizzabile in una o più regole.

Tuttavia ricondurre la categoria ad un modello è stato, ad esempio, possibile per le citazioni invertite.

All'interno della categoria citazioni invertite abbiamo distinto il caso in cui fosse invertita la collocazione della partizione rispetto al nome dell'atto (citazione invertita propriamente detta), dal caso in cui fossero invertiti, all'interno della partizione, i singoli elementi che la compongono (citazione con partizione invertita).

Per la formalizzazione di citazioni invertite propriamente dette, del tipo: "regio decreto 23 maggio 1924, n. 827, articolo 35", abbiamo redatto le regole seguenti.

New-citazione-invertita	New-citazione-invertita02	New-citazione-invertita03
TESTO- UNICO:vstart+DI?+DETX?+(DI SP-TRANS:vnome NOME- ATTO:vnome)? +[M_ALLCAT]+(DI+PX+DETX)?+(FONTE- COGNIZIONE+PX)?+[M- ATTO]+PUNCTX? ARTICOLO+[M-COORD- 1ART]+ (PUNCTX+DETX?+((M- ORDINALE- COOR]+COMMA+PUNCTX?) (COMMA+[M-COORD- 1COM]))?+(PUNCTX+ALINE A:valinea)?+(PUNCTX+DETX? +LETTERA+[M-COORD- 1LET])?+(PUNCTX+ALINEA:v alinea)?+(PUNCTX+DETX?+PA RTNUM+[M-COORD- 1NUM])?+[M-ALLCAT-END]	[M-NOME-ATTO]+ (AN-NUM- ATTO:vabbrevdata (DATE:vdata+(PUNCTX)?+PAR TNUM+(NUMBER:vnum))+PU NCTX?+ ARTICOLO+[M- COORD-1ART]+ (PUNCTX+DETX?+((M- ORDINALE- COOR]+COMMA+PUNCTX?) (COMMA+[M-COORD- 1COM]))?+(PUNCTX+ALINE A:valinea)?+ (PUNCTX+DETX?+LETTERA +[M-COORD- 1LET])?+(PUNCTX+ALINEA:v alinea)?+(PUNCTX+DETX?+PA RTNUM+[M-COORD- 1NUM])?+[M-ALLCAT-END]	CODICI:vcodici+PUNCTX? + ARTICOLO+[M-COORD- 1ART]+ (PUNCTX+DETX?+((M- ORDINALE- COOR]+COMMA+PUNCTX?) (COMMA+[M-COORD- 1COM]))?+(PUNCTX+ALINE A:valinea)?+ (PUNCTX+DETX?+LETTERA +[M-COORD- 1LET])?+(PUNCTX+ALINEA:v alinea)?+ (PUNCTX+DETX?+PARTNUM +[M-COORD-1NUM])?+[M- ALLCAT-END]

Abbiamo deciso di individuare i casi in cui il nome della partizione compaia solo una volta e non ci siamo occupati delle partizioni superiori all'articolo.

Per il caso delle citazioni con partizioni invertite le regole create sono più di una, perché molteplici sono le forme che abbiamo verificato dopo l'analisi del campione e per le quali non siamo stati in grado di costruire un unico modello.

Procedendo con lo stesso sistema casistico, sopra individuato ed a scopo meramente esemplificativo, abbiamo redatto anche regole per individuare singoli casi di citazione riconducibili alle categorie: citazione con elementi mancanti o pleonastici, citazione con

abbreviazioni e scioglimenti scorretti e citazioni *storiche*.

Nella classificazione proposta delle citazioni malformate la categoria delle citazioni storiche è stata individuata in base ad un criterio temporale difforme o incompatibile con il criterio linguistico strutturale, seguito per individuare le altre categorie; tale anomalia è comprovata dal fatto che con una sola regola abbiamo fornito un esempio di formalizzazione sia delle citazioni con abbreviazioni e scioglimenti scorretti sia di quelle storiche. Si tratta quindi di una imperfetta categorizzazione che utilizza criteri diversi; la conserviamo, nella consapevolezza che casi di citazioni malformate possono essere quantitativamente più estesi nelle citazioni storiche, piuttosto che nelle citazioni recenti.

Non escludiamo la possibilità di definire in seguito modelli corrispondenti a tali categorie e di ricorrere così alla compilazione di regole casistiche, la cui efficacia è sempre aleatoria o comunque limitata e il cui numero tende a moltiplicarsi. Probabilmente l'analisi di un corpus più vasto potrà consentire d'individuare, per così dire *a posteriori*, modelli generali anche per alcuni tipi di citazioni malformate.

Le prime analisi svolte su leggi emanate negli anni '90 e facenti parte del corpus normativo selezionato, confermano, per ora, che, grazie ai modelli di citazione regolare appena descritti, è possibile individuare ed estrarre oltre il 95% dei riferimenti normativi testuali espliciti, conformi alle regole della tecnica legislativa.

Schema riassuntivo delle tipologie di citazioni formalizzate:

Citazione Benformata	Atto	Normale	Sstandard Testi unici Senza numero
		Non paradigmatica Semplificata	Semplificata 1 Semplificata 2
		Multipla	
	Partizione	Sovrapartizioni	
		Articolo ed elementi inferiori	
		Annessi	
		Multipla	
Citazione Malformata modellizzabile	Citazione invertita		
	Partizione invertita		
Citazione malformata casistica	Cit. storica		
	Cit. con errori ortografici		
	Cit. con abbreviazioni e scioglimenti scorretti		
	Cit. con elementi mancanti o pleonastici		

Bibliografia:

S. Baroncelli, S. Faro, *Tecnica legislativa e legislazione regionale: l'esperienza delle regioni Toscana, Emilia Romagna e Lombardia*, in *Iter Legis*, 1998, pp. 173-213;

C. Biagioli, E. Marinai, P. Mercatali, *Documento normativo: note esplicative del DTD per i documenti normativi*, in *Informatica e diritto*, 1, 2000, pp. 55-106;

C. Biagioli, Ipotesi di modello descrittivo del testo legislativo per l'accesso in rete a informazioni giuridiche, <ipotesi2001.html> contribution to the Feasibility Study for the Implementation of the "Access to the Law on the Net", 31 gennaio 2000, pp. 1-90, www.ittig.cnr.it;

A. Bolioli, L. Dini, P. Mercatali, F. Romano: *For the automated mark-up of Italian legislative texts in XML*, in atti del Convegno Jurix 2002, Londra 16-17 dicembre 2002;

A. Boer, R. Hoekstra, R. Winkels, T. van Engers, F. Willaert, Proposal for a Dutch Legal XML Standard. (PDF), in Proceedings of the EGOV 2002 Conference (DEXA 2002).

<http://www.celi.it/conversione.htm>

M.C. De Lorenzo, Modelli di novelle, in *Informatica e diritto*, 1, 2002.

G. Ferrari, *Introduzione alla linguistica computazionale* in http://apollo.vc.unipmn.it/~ling_gen/opening.htm;

A. Marchetti, F. Megale, E. Seta, F. Vitali, Marcatura XML degli atti normativi italiani. I DTD di Norma in rete, in *Informatica e diritto*, 1, 2001, pp. 123-148;

P. Mercatali., *Atti normativi: criteri per l'individuazione e la descrizione*, in www.arslegis.org;

P. Mercatali, Produzione legislativa, informatica e intelligenza artificiale. Sistema sperimentale per il drafting legislativo, in Mercatali P., Soda G., Tiscornia D., Progetti di intelligenza artificiale per la pubblica amministrazione, Franco Angeli, Milan 1996, pp. 84-98.

B. Mortara Garavelli, *Le parole della giustizia*, Einaudi, Turin, 2001;

R. Pagano, *Le direttive di tecnica legislativa in Europa*, Camera dei deputati, Rome, 1997;

R. Pagano, *Introduzione alla legistica. L'arte di fare leggi*, Giuffrè, Milan, 2001;

G.U. Rescigno, *L'atto normativo*, Giappichelli, Torino, 1998;

E. Roche, Y. Schabes (eds.), *Finite-State Language Processing*, The MIT Press, 1997.

F. Sabatini, *Analisi del linguaggio giuridico*, in D'Antonio M., *Corso di studi superiori legislativi 1988-1989*, CEDAM, Padua, 1990;

G. Sartor, *Riferimenti normativi e dinamica dei nessi normativi*, in *Il procedimento normativo regionale*, Cedam, Padua, 1996.

T.M. van Engers, R.A.W. Vanlerberghe, *The POWER light-version: Improving Legal Quality under Time Pressure*, in <http://lri.jur.uva.nl/~epower> (2002).