

LESSICO ISTITUZIONALE DELLA STRATEGIA “EUROPA 2020” Confronto interlinguistico tra italiano, portoghese europeo e portoghese brasiliano¹

DANIELA VELLUTINO¹, GIAN LUIGI DE ROSA²

¹UNIVERSITÀ DI SALERNO, ²UNIVERSITÀ DEL SALENTO

Abstract – “Europe 2020” is the European Union strategy for smart, sustainable and inclusive growth through economic, social and territorial policies. The European structural and investment funds (ESIFs) are the primary form of investment for EU countries. EU institutions have invented a range of terminological expressions to manage these funds. These terms are mostly Multi-Word Expressions (MWEs) referring to legal and economic principles, bodies and procedures. For the purpose of our study we created a bilingual corpus of Italian and Portuguese institutional texts from which to extract the terms of this specific domain of EU terminology. Our study aims to analyze the EU institutional lexicon and outline the characteristics of this specific knowledge domain. Furthermore, the Committee of the Regions (CoR) decided to work in partnership with the Outermost Regions (ORs) and newly-emerging third countries, including Brazil, to achieve the strategic goals of “Europe 2020”. Brazilian Portuguese includes lexical forms that are different from those in European Portuguese, but, according to this EU decision, Brazilian Portuguese should be used for trading between the EU and Brazil. Our study found examples of Portuguese lexical forms that are not consistent with Brazilian terminology. Therefore, we think that our study may be useful to develop Linguistic Linked Open Data resources with also the lexical variants of Brazilian Portuguese for knowledge management through the Semantic web.

Keywords: terminology; Lexicon-Grammar; Multi-Word Expressions; Computational Linguistics; Linguistic Linked Open Data

*L'Unione rispetta la diversità culturale,
religiosa e linguistica.*
(Articolo 22.

Carta dei diritti fondamentali dell'Unione europea)

1. Alcune considerazioni preliminari

Il lessico istituzionale dell'Unione europea è in continua evoluzione: ogni giorno sono prodotti testi di vario tipo che contengono termini e fraseologismi appositamente creati per denominare concetti, istituti giuridici, organismi, procedure, ma anche titoli di documenti, programmi, fondi di finanziamento e quant'altro necessita alle complesse attività giuridico-amministrative ed informative. Tali testi sono prodotti in una delle tre lingue procedurali – inglese, francese e tedesco – e poi tradotti nelle 24 lingue ufficiali dell'Unione europea.

¹ I paragrafi e sottoparagrafi 1; 2; 3, 3.1; 6 sono di Daniela Vellutino; Il paragrafo 5 è di Gian Luigi De Rosa e il paragrafo 4 è stato scritto congiuntamente.

Pertanto pur essendo *de facto* traduzioni, per il principio della pari dignità linguistica,² che istituisce nella UE il multilinguismo integrale, sono *de jure* testi ufficiali originali.

Il contributo presenta i primi risultati di un'indagine sui testi scritti delle versioni in italiano, portoghese europeo e portoghese brasiliano (d'ora in poi PE e PB) volta ad esplorare i processi di variazione interlinguistica e intralinguistica a partire dall'analisi dei termini in forma di *espressioni complesse*³ o *lessemi complessi* per adottare la terminologia di Tullio De Mauro.

I termini su cui abbiamo concentrato la nostra attenzione sono quelli costituiti da lessemi complessi o polirematiche così spiegati da Tullio De Mauro nel *Grande dizionario italiano dell'uso* (De Mauro 1999):

Emergono, tra le *iuncturae*, quelle in cui il valore delle singole parole si trasvaluta e il nesso nella sua globalità appare con un imprevedibile valore nuovo rispetto ai valori dei singoli componenti. (De Mauro 1999, p. viii)

Le polirematiche, specialmente quando sono formazioni neologiche che si attestano per la prima volta nei testi dei documenti programmatici dell'Unione europea nelle versioni delle lingue ufficiali, non sono facilmente riconoscibili dagli strumenti di estrazione terminologica automatica. Spesso questi strumenti non riconoscono come termini di dominio lessemi complessi come ad esempio *specializzazione intelligente* oppure *programma operativo*. Pertanto, è necessario l'intervento dell'esperto di dominio per la loro individuazione.

Per il nostro studio i lessemi complessi sono stati estratti manualmente con l'aiuto di una esperta di dominio⁴ analizzando i testi istituzionali - legali e informativi - degli organismi dell'Unione europea sulla strategia "Europa2020". Sono stati ricercati i testi che riguardano il tema "politica di coesione", vale a dire la programmazione dei fondi strutturali e d'investimento nel periodo 2014-2020 destinati a rafforzare le imprese, incrementare l'occupazione e migliorare il tessuto sociale.

La "politica di coesione" non è tra i domini terminologici di Eurovoc,⁵ il thesaurus multilingue e pluridisciplinare, su cui si fonda anche IATE,⁶ la banca dati terminologica

² Nell'Unione europea tutte lingue degli Stati membri hanno pari dignità, tale principio è affermato sia nell'articolo 22 della *Carta dei diritti fondamentali dell'Unione europea*, in cui si riconosce espressamente il multilinguismo (Gazzetta ufficiale delle Comunità europee C 364/1 del 18.12.2000), sia negli articoli 24, 55 e 342 del *Trattato sul Funzionamento dell'Unione europea* 'TFUE' (Gazzetta ufficiale dell'Unione europea C 83/13 del 30.3.2010).

³ Il termine *espressione complessa* è qui usato come traduttore dell'anglicismo *MultiWord Expressions* (MWE), vale a dire una locuzione che esprime tanti fenomeni lessicali, quali: collocazioni, polirematiche, espressioni idiomatiche, fraseologie (Sag *et al* 2002).

⁴ Gli autori ringraziano per il suo contributo Simona Sabatino, esperta di dominio sulla comunicazione dei fondi strutturali e d'investimento europei.

⁵ EuroVoc è il thesaurus multilingue e pluridisciplinare che comprende la terminologia dei settori delle attività parlamentari dell'Unione europea. Contiene termini in 24 lingue dell'UE (bulgaro, ceco, croato, danese, estone, finlandese, francese, greco, inglese, irlandese, italiano, lettone, lituano, maltese, neerlandese, polacco, portoghese europeo, rumeno, slovacco, sloveno, spagnolo, svedese, tedesco e ungherese), oltre che in albanese, in serbo e nella lingua dell'ex Repubblica jugoslava di Macedonia (FYROM): <http://eurovoc.europa.eu/drupal/?q=it> (30.03.2016).

⁶ La banca dati IATE è gestita dall'Unità di Coordinamento Terminologico, TermCoord, che opera attraverso il contributo di traduttori professionisti con profili terminologici specializzati nelle 11 istituzioni UE: Commissione europea, Parlamento europeo, Consiglio dell'Unione europea, Corte di giustizia europea, Corte dei conti europea, Comitato economico e sociale europeo, Comitato delle Regioni, Banca centrale

multilingue interistituzionale e interattiva dell’Unione europea. Eppure la “politica di coesione” è un tema di primaria importanza per i cittadini europei perché la sua programmazione impegna il 34% del bilancio UE, pari a 325 miliardi di euro investiti per lo sviluppo delle Regioni dei 28 Stati membri.

Per il nostro studio abbiamo ricercato nelle lingue indagate i termini in forma di lessemi complessi pertinenti questo tema con le loro varianti lessicali, vale a dire sinonimi o più di frequente forme abbreviate, quali acronimi e sigle, presenti nei differenti tipi di testo prodotti per i diversi contesti pragmatici dell’informazione istituzionale e della comunicazione pubblica (Vellutino 2014).⁷

I termini sono stati analizzati descrivendo i tipi di composizione lessicale al fine di individuare analogie nelle strutture combinatorie lessicali delle lingue indagate.

Il riconoscimento e l’analisi dei lessemi complessi potrà essere utile per lo sviluppo di algoritmi per il trattamento automatico del linguaggio (*Natural Language Processing*, NLP). Infatti, la descrizione della formazione dei termini nelle lingue neolatine indagate consente di dare istruzioni per la codifica dei linguaggi di marcatura e per i software per il trattamento automatico del linguaggio, tra i quali anche i sistemi di traduzione assistita basati sulle regole di combinazione lessicale.

Questa nostra indagine è parte di uno studio finalizzato allo sviluppo di risorse linguistiche in Italiano/PE/PB, in formato *Linguistic Linked Open Data*,⁸ utile allo sviluppo di software per il trattamento automatico del linguaggio.

2. Note sulle fonti dei testi istituzionali della strategia “Europa 2020”

Il termine istituzionale “Europa 2020” denomina la strategia politica proposta dalla Commissione europea nel 2010 per sostenere l’occupazione, la produttività e la coesione sociale negli Stati membri dell’Unione.

La sua prima attestazione è nella Comunicazione della Commissione del 3 marzo 2010 “Europa 2020: Una strategia per una crescita intelligente, sostenibile e inclusiva” [COM(2010) 202 def. - Non pubblicata nella Gazzetta ufficiale]. Questo documento istituzionale è il primo atto d’indirizzo a cui sono collegati gli atti⁹ redatti nelle lingue indagate, ma non sempre tutti i testi che riguardano questo argomento sono tradotti in tutte le lingue ufficiali dell’Unione.

europea, Banca degli investimenti europea, Centro di traduzione per gli organismi dell’Unione europea, Ufficio delle Pubblicazioni. <http://iate.europa.eu/SearchByQueryLoad.do?method=load> (30.03.2016).

⁷ Il design del corpus segue il modello di classificazione dei testi istituzionali elaborato da Daniela Vellutino (Vellutino 2014) che distingue i testi istituzionali a seconda delle attività d’informazione e comunicazione delle amministrazioni pubbliche in relazione agli obblighi di pubblicazione definiti dalle norme italiane ed europee in materia.

⁸ *Linguistic Linked Open Data* (LLOD) sono i dati aperti che rispettano i 4 principi base proposti da Tim Berners-Lee in *Linked Data: Principles and State of the Art* <https://www.w3.org/DesignIssues/LinkedData.html>.

⁹ Conclusioni del Consiglio europeo di Bruxelles del 25 e 26 marzo 2010”; “Raccomandazione 2010/410/UE del Consiglio del 13 luglio 2010, sugli orientamenti di massima per le politiche economiche degli Stati membri e dell’Unione” [Gazzetta ufficiale L 191 del 23.7.2010]; “Decisione 2010/707/UE del Consiglio del 21 ottobre 2010 sugli orientamenti per le politiche degli Stati membri a favore dell’occupazione” [Gazzetta ufficiale L 308 del 24.11.2010].

Si è rilevato, infatti, che non tutti i documenti istituzionali che riguardano questo argomento sono accessibili nelle lingue ufficiali dell'Unione e sui siti delle fonti istituzionali del dominio web.eu dove un cittadino europeo potrebbe accedere, oltre a cercarli sul portale specializzato del diritto europeo Eur-Lex. Nonostante quanto affermato a riguardo della politica linguistica dell'Unione europea nel *Trattato sul Funzionamento dell'Unione europea* 'TFUE', nella *Carta dei diritti fondamentali dell'Unione europea*; poi ribadito da una sentenza della Corte di Giustizia europea che riguarda in particolare le traduzioni in lingua italiana.¹⁰

La stessa principale fonte istituzionale per i documenti d'indirizzo della "politica di coesione", la *Direzione Generale della Politica regionale e urbana della Commissione* (DG REGIO), sul suo sito web istituzionale non ha pubblicato tutti i testi legali sull'approvazione della strategia "Europa 2020" in tutte le lingue ufficiali dell'Unione.

La DG REGIO, oltre a gestire le risorse finanziarie, ha tra le sue competenze anche il coordinamento degli interventi per le Regioni Ultraperiferiche (RU), i territori oltre confine di alcuni Stati europei, tra i quali anche i territori insulari del Portogallo.

Infatti, non tutti i testi istituzionali sono tradotti dai dipartimenti di lingua delle istituzioni e degli organismi UE e dal Centro di Traduzione dell'Unione europea (CdT). In particolare, spesso non sono tradotti quei testi che riportano informazioni specialistiche utili per la promozione degli scambi commerciali e per la condivisione di conoscenze con i Paesi terzi, quali i principali Paesi emergenti (ossia il Brasile e il Sud Africa). Eppure questi scambi commerciali potrebbero essere utili per la "crescita", in considerazione del fatto che l'UE trae vantaggio dalle assidue relazioni che le Regioni Ultraperiferiche (RU) intrattengono con i principali Paesi emergenti di lingua portoghese con le sue varietà diatopiche.¹¹

Il nostro studio parte dall'individuazione e dall'estrazione dei lessemi complessi dai documenti d'indirizzo programmatico prodotti dagli organismi unionali e dai numerosi testi istituzionali delle versioni nelle lingue indagate in cui sono presenti termini e fraseologie che denominano concetti relativi a questo specifico argomento.

Riteniamo, infatti, che la divulgazione della terminologia sulla politica di coesione e la corretta traduzione dei testi istituzionali abbia una rilevante importanza strategica per il raggiungimento dell'obiettivo della crescita intelligente, sostenibile e inclusiva della strategia "Europa 2020".

3. Prime note sulla composizione dei lessemi complessi del lessico istituzionale dell'Unione europea

Negli ultimi anni è aumentato l'interesse per lo studio dei lessemi complessi nelle lingue comuni e nelle terminologie specialistiche (Voghera 2004).

Per la nostra indagine abbiamo come riferimento metodologico gli studi di Annibale Elia, che ha analizzato diversi tipi di polirematiche seguendo l'approccio Lessico-Grammatica (LG)¹² finalizzato allo sviluppo di dizionari elettronici di dominio, vale a dire

¹⁰ La sentenza della Corte di Giustizia dell'UE sul sito InfoCuria <http://curia.europa.eu/juris/documents.jsf?num=C-566/10> (25.07.2016).

¹¹ L'interesse per incentivare gli scambi commerciali è stato espresso dal Comitato delle Regioni nel parere "Le regioni ultraperiferiche dell'UE alla luce della strategia Europa 2020" (2013/C 62/13).

¹² L'apparato teorico-metodologico del Lessico-Grammatica (LG) si sviluppa a metà degli anni '70 con gli studi di Maurice Gross in Francia e di Annibale Elia in Italia. Per LG il lessico non è separabile dalla sintassi

risorse lessicografiche formalizzate in basi di dati utili per gli applicativi di trattamento automatico del linguaggio (NLP).

Gli studi di Annibale Elia sulle polirematiche in lingua italiana a partire dagli anni ottanta del '900 sono precedenti agli studi attualmente molto citati nella letteratura scientifica (Sag *et al.* 2002) che riportano descrizioni utili per tipologizzare le strutture compositive dei lessemi complessi.

Per gli strumenti di estrazione terminologica automatica le polirematiche possono essere nel contempo un'opportunità ed una criticità. Una criticità perché tali strumenti sono per lo più fondati su misure statistiche¹³ per rilevare le concordanze e la coesione tra gli elementi lessicali che le compongono. Se le concordanze non hanno rilevanza statistica all'interno dei testi istituzionali di dominio, difficilmente i lessemi complessi possono essere riconosciuti automaticamente in tutte le lingue delle versioni dei testi istituzionali unionali. Invece, se individuate attraverso risorse lessicografiche formalizzate, come sostiene Elia,¹⁴ quali sono per l'appunto i dizionari elettronici, le polirematiche possono diventare un'opportunità per gli strumenti di trattamento automatico delle lingue perché facilitano la disambiguazione delle parole comuni nei lessemi complessi. Inoltre, una volta riconosciute come sequenze lessicali con valore terminologico, è possibile creare vocabolari controllati¹⁵ di dominio che consentono i processi di indicizzazione, di recupero dell'informazione e d'interoperabilità semantica.

Da uno studio in corso partito dalla caratterizzazione della varietà dell'italiano istituzionale (Vellutino *et al.* 2012, p. 553) si è osservato che ciò che maggiormente caratterizza la formazione del lessico istituzionale dell'Unione europea è proprio la costruzione di lessemi complessi in cui è presente in posizione testa un nome della lingua comune – ad esempio *programma, complemento, pista, asse, accordo* – seguito da altri elementi lessicali con variabile forza di coesione. Sono proprio questi elementi che seguono il nome testa del sintagma che determinano il significato del termine e, dunque, svolgono la funzione di arricchimento semantico. Le neoformazioni del lessico istituzionale europeo, però, sono il prodotto di neologia secondaria (Sager 1990, pp. 80-81), vale a dire che i termini in forma di lessemi complessi delle lingue ufficiali dell'Unione europea sono repliche della lingua di partenza dei testi istituzionali, il cosiddetto *euro-english* (Gardner 2013), formate attraverso la sostituzione degli anglicismi con gli equivalenti delle lingue di arrivo.

Dal nostro studio, infatti, si è rilevato un analogo comportamento linguistico anche nel PE. Ciò è dovuto alla comune matrice neolatina che regola la formazione dei calchi dalla lingua di partenza dei testi istituzionali nelle lingue ufficiali dell'Unione europea che sono sì originali, ma traduzioni *de facto*. Il calco di traduzione (Bombi 2005, pp. 53-54) è uno dei principali meccanismi di formazione neologica del lessico istituzionale dell'Unione europea e delle terminologie specialistiche spesso contenute nei testi istituzionali. Nei testi indagati,

perché ogni elemento lessicale segue regole di grammatica per le sue combinazioni di co-occorrenza, di restrizione e selezione in relazione alle proprietà degli altri elementi lessicali del contesto frastico. In base a tali studi sono state sviluppate risorse linguistiche computazionali quali Tavole Lessico-Grammaticali, dizionari elettronici e grammatiche locali.

¹³ Principali misure di rilevanza statistica: *Term frequency-inverse document frequency* (tf-idf) (Salton, Buckley 1988); *Mutual Information* (Church, Hanks 1990); *Log-likelihood ratio* (Dunning 1993); *C-value/NC-value* (Fratzi, Ananiadou 1999).

¹⁴ I dizionari elettronici sono basi di dati lessicali formalizzate, strutturate omogeneamente, in cui le caratteristiche morfo-grammaticali, sintattiche e semantiche per il settore d'uso sono indicate da etichette alfanumeriche univoche e non ambigue, tali da consentirne la lettura e la comprensione da parte di un elaboratore elettronico (Elia 1995).

¹⁵ Vocabolario controllato è uno schema di classificazione dei termini utile per i processi di indicizzazione e di ricerca e recupero delle informazioni dei e nei documenti.

però, si è osservato che le neoformazioni sono calchi adattati alle esigenze morfosintattiche delle lingue neolatine di arrivo. Pertanto la struttura compositiva interna del lessema complesso inglese costituito dalla sequenza Nome+Nome (N+N) spesso è resa utilizzando il nesso di specificazione dato dalla preposizione seguita dal nome, componendo la sequenza N+prep+N, come i lessemi composti dai nomi comuni *programma/programa, pista/pista, azione/ação, accordo/acordo* che ricorrono in diversi tipi di sequenze lessicali; nei calchi dall'*euro-english* di dominio *action programme, partnership agreement, audit trail* sono resi con il nesso di specificazione: *programma di azione/ programa de ação; accordo di partenariato/acordo de parceria; pista di controllo/ pista de controle*.

Nelle lingue indagate le composizioni più ricorrenti hanno strutture sintagmatiche composte dai costituenti lessicali:

Nome+Preposizione+Nome (N+prep+N): *programma di azione/programa de ação; complemento di programmazione/complemento de programação; pista di controllo/pista de controle*

Nome+Nome (N+N): *Programma Erasmus/Programa Erasmus; azione chiave/ ação-chave*

Nome+Aggettivo (N+AGG): *Programma operativo/Programa operacional; asse prioritario/eixos prioritários*

Tuttavia, se prendiamo ad esempio la parola semplice *accordo/acordo*, vediamo che questo lessema appartiene sia alla lingua comune che a diverse terminologie specialistiche. Ciò crea difficoltà per la sua disambiguazione attraverso gli strumenti per il trattamento automatico del linguaggio.

Nel lessico istituzionale unionale il lemma *accordo/acordo* ricorre spesso in posizione di testa nei lessemi complessi in cui i costituenti che seguono possono disambiguarlo, ad esempio nei tipi delle strutture interne dei lessemi complessi:

N+Prep+N: *accordo di cooperazione / acordo de cooperação*

N+N: *accordo quadro / acordo-quadro*

N+AGG: *accordo transattivo / N+Prep+N: acordo de transação*

Nel nostro studio, sebbene sia stato condotto su corpora ristretti, abbiamo osservato che diversi tipi di lessemi complessi sono presenti nelle lingue indagate. In particolare, sono stati rilevati tipi analoghi che appaiono interessanti per gli esiti computazionali di disambiguazione.

I casi più frequenti sono i lessemi complessi con struttura sintagmatica fissa determinata da una rigidità morfosintattica che non ammette variazioni interne come ad esempio la pluralizzazione (esempi tipo A). Questo è il comportamento linguistico dei nomi propri d'istituzione, dei titoli dei documenti, che, però, proprio in virtù della loro lessicalizzazione, ammettono la creazione di varianti lessicali quali le abbreviazioni in forma di acronimi o sigle.

Un altro tipo di lessemi complessi individuati consente, invece, la pluralizzazione di tutti gli elementi (esempi tipo B). Infine, si è rilevato ancora un altro tipo di sequenze lessicali che consente l'inclusione di altri elementi lessicali nella struttura sintagmatica (esempi tipo D).

- A) Lessemi complessi a struttura fissa**
Nomi delle istituzioni
Unione europea **Unioni europee* / UE
União europeia * *Uniões europeias* /UE

È da rilevare che l’acronimo UE è diventato un internazionalismo che rappresenta una variante per frequenza quasi concorrente dell’equivalente acronimo inglese EU da *European Union*.

- Titoli dei documenti**
EUROPA 2020 Strategia per una crescita intelligente, sostenibile ed inclusiva
EUROPA 2020 Estratégia para um crescimento inteligente, sustentável e inclusivo

- B) Lessemi complessi che consentono la pluralizzazione dei costituenti del sintagma nominale:**
Bollettini Ufficiali / *Boletins Oficiais*

- C) Lessemi complessi che consentono la pluralizzazione solo del nome testa:**
Comitati di sorveglianza / *Comissões de Acompanhamento*

- D) Lessemi complessi che consentono l’inclusione di altri elementi lessicali nella struttura sintagmatica:**
Sviluppo (rurale) locale / *Desenvolvimento (rural) local*

3.1. Studio pilota

Per questo primo studio esplorativo siamo partiti dall’individuazione dei testi istituzionali rilevanti nelle lingue indagate per comunicare la strategia “Europa 2020”, vale a dire quelli pertinenti la politica di coesione per il ciclo di programmazione dei fondi strutturali 2014-2020, al fine di creare il repertorio di testi istituzionali unionali da cui estrarre i termini attraverso una lettura diretta condotta con l’esperta di dominio.

Questa fase è stata preliminare allo sviluppo di corpora paralleli bilanciati che saranno in seguito analizzati anche attraverso applicativi per il trattamento automatico della lingua che abbiano moduli operativi per le lingue osservate.

I testi sono stati collezionati seguendo il modello di classificazione dei tipi di testo istituzionali basato sulle finalità pragmatiche delle attività d’informazione istituzionale e di comunicazione pubblica. Pertanto, al fine di costruire corpora paralleli sono stati individuati e collezionati tipi di testo per la pubblicità legale¹⁶ e tipi di testo informativi per la comunicazione pubblica.

Testi per la Pubblicità Legale	Tokens
Italiano	65205
Portoghese	75413

Tabella 1
 Dimensioni dei corpora della pubblicità legale.

Testi per la Pubblica utilità	Tokens
-------------------------------	--------

¹⁶ La pubblicità legale è una forma di comunicazione obbligatoria per far conoscere, attraverso appositi testi e mezzi di comunicazione, fatti, atti giuridici ed amministrativi specificati dalle norme italiane e dell’Unione europea in materia. A seguito di questa pubblicazione, tali fatti ed atti acquisiscono validità giuridica.

Italiano	25021
Portoghese	28499

Tabella 2
Dimensioni dei corpora della comunicazione pubblica.

Dai testi sono state estratte manualmente i lessemi complessi che riguardano il dominio ristretto “Politica di coesione 2014-2020”, creando lemmari per ogni lingua.

È stato così individuato un primo nucleo di terminologia istituzionale specialistica, costituito da 30 lessemi complessi che denominano organismi, titoli di documenti, attività del processo di programmazione, modalità operative.

Alcuni lessemi complessi sono formazioni neologiche: *Specializzazione intelligente / Especialização Inteligente* oppure *Concentrazione tematica*/ Concentração temática*; altre, come ad esempio *Quadro Strategico Comune (QCS) / Quadro Estratégico Comum (QEC)* e *Codice di Condotta Europeo / Código de Conduta Europeu* sono già presenti nei documenti di programmazione dei cicli precedenti 2007-2013 e 2000-2006, ma nei testi indagati hanno acquisito nuovi significati. Dunque, hanno subito un processo di risemantizzazione di cui è necessario tener conto.

Tali lessemi complessi, sebbene non siano statisticamente frequenti nei testi osservati, sono portatori di un significato specifico spesso giuridicamente vincolante. Ad esempio, il termine *Programma operativo/Programa operacional (PO)* nel lessico istituzionale della “Politica di coesione 2014-2020” ha un significato specifico definito nel Regolamento 1303/2013 (art. 96), differente rispetto alla definizione descritta nel Regolamento della precedente programmazione dei fondi strutturali 2007-2013. Anche altri termini, come ad esempio, *Asse prioritario/Eixo prioritário* (con variante monorematica *priorità/prioridade*); *Aiuti di stato/Auxílio de estado*; *Organismo intermedio/Organismo intermediário*, sono termini il cui significato è definito nella proposta di Regolamento (COM(2013) 246 final 2011/0276 (COD) e sono presenti sia nei testi della pubblicità legale che in quelli informativi per la comunicazione pubblica.

Anche questa considerazione sulla risemantizzazione dei termini del lessico istituzionale, dovuta alla loro origine negoziale all’interno delle commissioni parlamentari dalla cui attività si producono i regolamenti in cui sono presenti le loro definizioni, ci induce ad ipotizzare l’utilità di repertoriare i lessemi individuati come termini di questo specifico dominio “Politica di coesione 2014-2020” in un vocabolario controllato utilizzabile dalla banca dati terminologica multilingue dell’Unione europea IATE.

4. Prime osservazioni per il confronto interlinguistico

L’indagine condotta ha rilevato che i termini individuati presentano strutture sintagmatiche ricorrenti sulla cui composizione e significazione è necessario avanzare alcune osservazioni per un confronto interlinguistico.

Sono stati individuati come termini i lessemi complessi che denominano organismi, attività del processo di programmazione e modalità operative. Ad esempio:

Autorità di gestione/ Autoridade de gestão*

Acronimo (internazionalismo) *AdG*

Gruppo europeo di cooperazione territoriale (GECT) / Agrupamentos europeus de cooperação territorial (AECT)

Assistenza tecnica / Assistência técnica

Cooperazione territoriale / Cooperação territorial

Aiuto di stato/Auxílio de estado

Investimenti territoriali integrati / Investimento territorial integrado*

In particolare, si è rilevato che i termini che hanno subito un processo di risemantizzazione hanno acquisito nuovi significati nella programmazione 2014-2020:

Contratto di partenariato / Contrato de parceria

Nella banca dati terminologica multilingue IATE, questa entrata è registrata insieme alla variante *Accordo di partenariato* che, invece, è assente nei testi della pubblicità legale e della comunicazione pubblica.

In alcuni testi italiani suggeriti dall'esperta di dominio, però, è attestato il termine *Accordo di partenariato* con l'acronimo AdP; i due termini, *Contratto di partenariato* e *Accordo di partenariato* non sono varianti che esprimono uno stesso concetto, se si tiene conto di quanto definito nel Regolamento (COM(2013) 246 final 2011/0276 (COD)).

Gruppi di azione locale / Grupos de ação local

Acronimo (internazionalismo) GAL/ GAL

Questo termine è già attestato nei precedenti cicli di programmazione dei fondi strutturali europei, ma nella politica di coesione 2014-2020 sta acquisendo un nuovo significato perché sono cambiate le sue finalità operative.

In assenza di una politica linguistica che definisce il processo di creazione dei termini dei vari domini istituzionali dell'Unione europea si producono rese traduttive con lessemi complessi non equivalenti tra le diverse lingue dell'Unione che sicuramente creano problemi ai sistemi di trattamento automatico del linguaggio NLP. Come ad esempio:

Community-led Local Development → Sviluppo locale di tipo partecipativo/ Desenvolvimento local orientado para a comunidade.

Un caso positivo di neoformazione lessicale è dato dal termine che si sta attestando nella forma plurale:

Integrated Territorial Investment → Investimenti territoriali integrati/ Investimentos territoriais integrados

Il termine si sta acronimizzando come internazionalismo *ITI / ITI* nelle altre lingue neolatine ed anche in inglese e in tedesco. Questa formazione consente di ottimizzare i processi di riconoscimento lessicale per il trattamento automatico del linguaggio NLP.

5. Note per il confronto tra PE e PB

Questo studio esplorativo sui tipi di composizione delle polirematiche istituzionali del PE nei diversi tipi di testo per la pubblicità legale e per la pubblica utilità consente di osservare

il processo di ristandardizzazione che interessa PE e PB e la complessità di relazioni che intercorrono tra PE e PB.

La maggior parte dei documenti analizzati ha evidenziato l'impiego nel PE di una serie di polirematiche, il cui uso si registra soltanto in parte nelle attività d'informazione istituzionale e comunicazione pubblica brasiliana.

Prima di addentrarci nella comparazione tra i diversi impieghi delle polirematiche in queste forme di comunicazione, è necessario evidenziare quanto la realtà policentrica del portoghese presenti due diasistemi, uno europeo, l'altro brasiliano, in cui si registrano processi di ristandardizzazione distinti e dissimili. Difatti, se il contesto portoghese presenta una realtà che potremmo considerare di tipo dilalico, in Brasile ci si trova dinanzi ad una realtà di tipo diglossico, nel senso che tra il PB standard e il PB neo-standard/varietà colte urbane si è creata una distanza incolmabile, così come si evince dalla figura 1, in cui è ritratta l'architettura del PB.

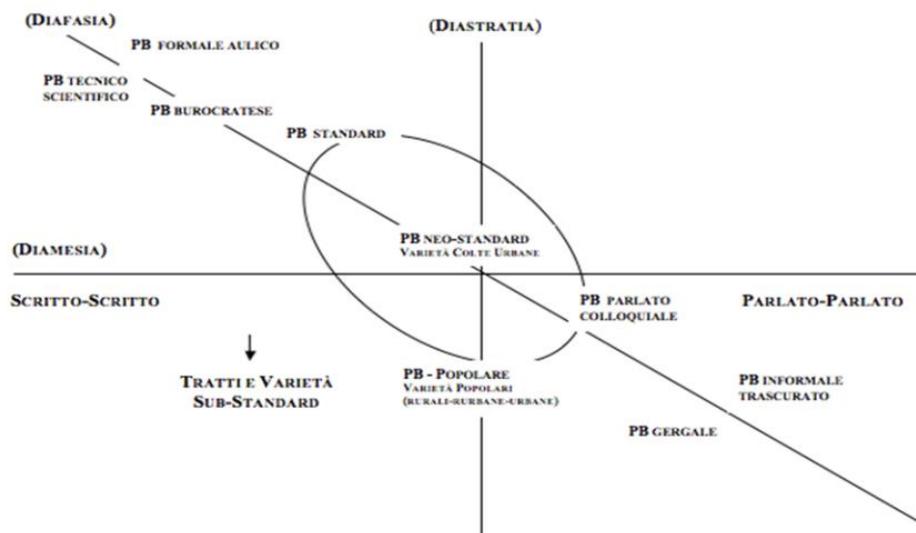


Figura 1
Architettura del PB.

I più recenti studi sociolinguistici sul PB ne hanno confermato, senz'ombra di dubbio, il tratto dell'eterogeneità, arrivando ad indicare il PB non solo come eterogeneo e variabile, ma anche come plurale e polarizzato.

In un certo senso, il PB si presenta come un diasistema linguistico fortemente polarizzato, ai cui poli ci sarebbero, da un lato, il sottosistema delle varietà colte urbane (PB neo-standard), varietà di prestigio occulto, e dall'altro il sottosistema delle varietà popolari, fortemente stigmatizzate, frutto, tra l'altro, della trasmissione irregolare che, a partire da quello che può considerarsi il primo tentativo di realizzare una politica linguistica riguardante la lingua portoghese, messa in atto dal marchese di Pombal (1757-1759), vide il portoghese imporsi su tutto il territorio brasiliano in detrimento delle lingue *gerais* (lingue indigene grammaticalizzate dai gesuiti).

La complessità di tale situazione linguistica si capisce meglio quando, ancora oggi, parlando di PB, si fa quasi sempre riferimento ad uno standard diverso dalle varietà colte urbane, riscontrabile e associato soprattutto alla modalità scritta della lingua. Tuttavia, già

confrontando le varietà colte urbane, impiegate in contesti di uso controllato e monitorato, con la norma scritta, si evidenziano notevoli divergenze, che diventano vere e proprie discrepanze quando il confronto avviene con le varietà popolari e sub-standard impiegate quotidianamente dalla quasi totalità della popolazione brasiliana (De Rosa 2008, 2009). Si può sintetizzare il tutto dicendo che la polarizzazione linguistica in Brasile vede da un lato le varietà colte urbane (varietà di prestigio occulto) e dall'altro le varietà popolari (varietà discriminate), mentre la varietà standard (varietà di prestigio manifesto) – divergente, altamente codificata ed appresa solo mediante l'istruzione formale – viene usata solo ed esclusivamente per scopi formali per determinati contesti pragmatico comunicativi con i relativi tipi di testo scritti, tra cui i documenti scritti della comunicazione istituzionale, così come attesta il Manual de Redação da Presidência da República del 2002:

(O)s textos oficiais: por seu caráter impessoal, por sua finalidade de informar com o máximo de clareza e concisão, eles requerem o uso do padrão culto da língua. Há consenso de que o padrão culto é aquele em que a) se observam as regras da gramática formal, e b) se emprega um vocabulário comum ao conjunto dos usuários do idioma. É importante ressaltar que a obrigatoriedade do uso do padrão culto na redação oficial decorre do fato de que ele está acima das diferenças lexicais, morfológicas ou sintáticas regionais, dos modismos vocabulares, das idiosincrasias lingüísticas, permitindo, por essa razão, que se atinja a pretendida compreensão por todos os cidadãos.

Pertanto, se la comunicazione quotidiana in Brasile si realizza principalmente mediante le varietà colte urbane (PB neo-standard), con elevata permeazione con le varietà popolari in un continuum rurale-rurbano-urbano (cfr. figura 1), mentre il PB standard – divergente e altamente codificato – viene usato solo per scopi formali e nella modalità di lingua scritta, si comprende che la conseguenza di tutto ciò è che le attività d'informazione istituzionale e la comunicazione pubblica, per essere intellegibili da parte della cittadinanza, dovrebbero impiegare la varietà neo-standard e non lo standard.

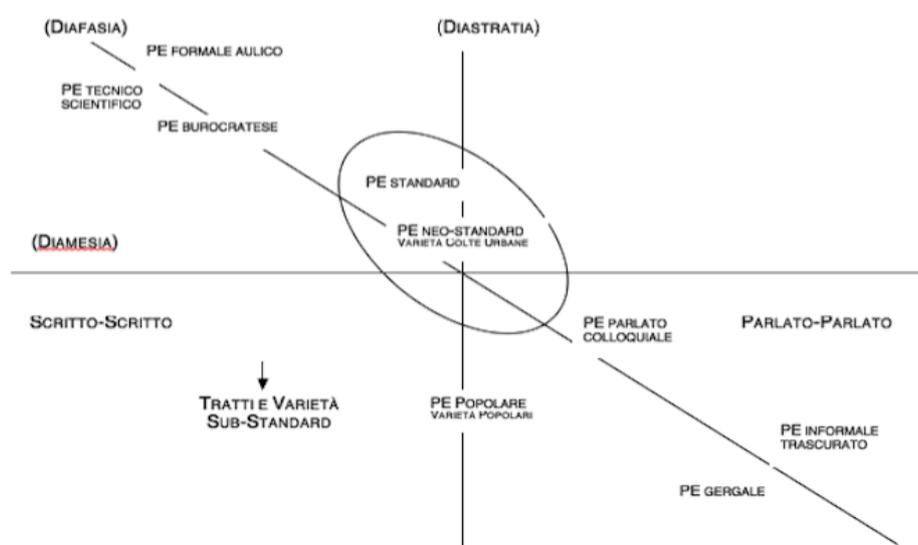


Figura 2
Architettura del PE.

Appurata tale situazione di fatto si è cercato di comprendere se quelle polirematiche, il cui impiego è attestato nella comunicazione ufficiale portoghese scritta, potessero essere utilizzate nei tipi e nei generi testuali che rimandano alla comunicazione ufficiale brasiliana.

Premettendo che si tratta di un primo approccio all'argomento e che saranno necessari ulteriori indagini e ricerche, emerge che le divergenze di tipo lessicale con il PE rivelano sacche di opacità che non permetterebbero la totale sovrapposizione delle due varietà istituzionali della lingua portoghese.

Difatti, entrando nello specifico delle polirematiche evidenziate, se nel caso della traduzione di *Accordo di Partenariato* avremmo due varianti di tipo fono-grafico (cambia l'accento), che non inficiano l'acronimo APE, ma che funzionano in entrambi i contesti istituzionali: *Acordo de parceria económica* (PE) / *econômica* (PB).

La situazione cambia nel caso di *Condicionalidade Macroeconómica* (PE) / *Macroeconômica* (PB), in cui avremmo due varianti che presentano caratteristiche totalmente diverse, essendo poco o per nulla diffusa tale polirematica nel contesto brasiliano.

Allo stesso modo, l'uso della polirematica *Programa operacional plurifundo* non è attestata in Brasile per il fatto di essere un termine istituzionale in uso nei programmi dell'UE. Anche nel caso di *Auxílio Estatal* (PE) / *Auxílio Estadual* (PB) ci si trova dinanzi a una coppia di varianti che riguardano il secondo elemento della polirematica. Difatti, nella polirematica brasiliana si passa da *Estatal* a *Estadual*, rendendo parzialmente opaca la ricezione della polirematica portoghese nel contesto brasiliano.

Diverso il caso di *Autorità di certificazione* polirematica che troviamo tradotta in PE con *Entidade de certificação*, mentre in Brasile la polirematica equivalente è *Autoridade de certificação*, che mostra nuove crepe nella possibilità di una sovrapposizione secca tra le due varietà istituzionali.

Quanto a *Governance multilivello*, la situazione è abbastanza diversificata tra PE e PB. In PB il traduttore riconosciuto è *Governança multinível*, visto che *Governance* trova in *Governança* il suo equivalente traduttivo più naturale, risultando ampiamente accettato. Per il PE la situazione è leggermente diversa, visto che, non esiste unanimità sul traduttore di *Governance*. A comprovare tale situazione c'è anche un'interrogazione parlamentare pubblicata nel 2002 sulla Gazzetta ufficiale dell'Unione europea (versione originale in portoghese), in cui si metteva in evidenza la problematicità dell'alternanza tra *Governança* e *Governança* come traduttori dell'anglicismo *Governance*.

INTERROGAZIONE SCRITTA E-2103/02 di Paulo Casaca (PSE) alla Commissione (16 luglio 2002)

Oggetto: Persistenza dell'«Ancien Régime» nella lingua portoghese

Tramite la risposta E-1616/02 (GU C 92 E del 17.4.2003, pag. 71), la Commissione mi ha comunicato che «qualsiasi connotazione negativa associata al termine (governança) risale all'Ancien Régime «ed è divenuta quindi obsoleta».

Vorrei far rilevare alla Commissione che, oltre ai dizionari citati e alla mia personale sensazione, esistono vari altri esempi della palese persistenza di questo spirito da «Ancien Régime».

L'espressione «governance» è stata tradotta, ad esempio, dalla Commissione europea con «governança» nel documento «Um Livro branco sobre a governança europeia - Aprofundar a democracia na União europeia» (SEC(2000) 1547/7 def. dell'11.10.2000) e, sempre nel luglio dell'anno scorso, è stato tradotto con «boa governança» nel dizionario elettronico della Commissione europea. La base terminologica di dati del Consiglio è inoltre un altro degli esempi di questa persistenza dell'«Ancien Régime», visto che utilizza altresì il termine «governança» e lo consacra in uno dei gruppi di lavoro della Convenzione sul futuro dell'Europa.

Risposta data dal sig. Kinnock a nome della Commissione (11 settembre 2002):

Il programma di lavoro per la preparazione del Libro bianco sulla governance europea adottato dalla Commissione nell'ottobre 2000 (SEC(2000) 1547/7 def.), a cui si riferisce l'onorevole parlamentare, non è il Libro bianco vero e proprio (COM(2001) 428 def. — GU C 287 del 12.10.20), ma solo un documento interno preparatorio inteso a riunire in un unico contesto

istituzionale il gruppo che si occupa della governance e i vari gruppi interservizi che collaborano con esso.

Il documento rivolge, fra l'altro, particolare attenzione ai problemi di traduzione e di comunicazione che il termine *governance* potrebbe porre nelle lingue dell'Unione.

In alcuni casi, dopo questo esame approfondito si sono cambiati i termini adottati inizialmente nelle traduzioni, abbandonando ad esempio il termine *governança*, con cui era stata tradotta *governance*, per adottare la traduzione che figura nel Libro bianco. Come si è già spiegato nella risposta data dalla Commissione all'interrogazione precedente dell'onorevole parlamentare (Interrogazione scritta E-1616/02), nella scelta definitiva si sono evitati deliberatamente i termini che potrebbero essere confusi con «governo» o con «attività governativa».

In termini di percezione, le ragioni dell'interrogazione parlamentare hanno origine nella percezione negativa del termine *governança* (e la sua conseguente percezione stigmatizzata), mettendone in evidenza anche i motivi. Tuttavia, anche se la Commissione si produce in una risposta favorevole all'impiego di *governança*, il risultato venuto fuori dalla nostra ricerca ha evidenziato una situazione diversa. Difatti, se per il PB è ampiamente accettato l'uso della polirematica *Governança multinível*, per il PE, nei testi istituzionali scritti è possibile l'alternanza, anche se nel nostro corpus è stato rilevato esclusivamente la polirematica *Governança a vários níveis*.

Anche in questo caso, come nel precedente, in Brasile si impiega una polirematica diversa che mostra crepe nella possibilità di una sovrapposizione secca tra le due varietà istituzionali.

Per terminare risulta ancora più interessante il caso di *Sviluppo locale di tipo partecipativo*, il cui equivalente in PE è la polirematica *Desenvolvimento Local Orientado para a Comunidade*, presente anche con l'acronimo DLOC, mentre in Brasile risulta maggiormente usata la polirematica *Desenvolvimento local participativo*.

Nonostante ciò, anche in Portogallo si registra l'uso dell'aggettivo "partecipativo" nella comunicazione pubblica, come è stato fatto per la campagna *orçamento participativo* del Comune di Lisbona, realizzata nell'autunno del 2013 (www.lisboaparticipa.pt). Tuttavia, è necessario evidenziare, in termini di significato, il fatto che *Desenvolvimento Participativo* veicola (inform)azioni di tipo bi-direzionale TOP>DOWN e DOWN>TOP, mentre *Desenvolvimento Local Orientado para a Comunidade* veicola (inform)azioni esclusivamente di tipo uni-direzionale: TOP>DOWN.

6. Prime conclusioni e prospettive future

Da queste prime evidenze empiriche possiamo rilevare che la varietà della lingua istituzionale osservata nelle strutture compositive dell'italiano è riscontrabile anche nel PE. Per il dominio terminologico osservato entrambe le lingue neolatine seguono equivalenti comportamenti nella formazione dei termini in forma di lessemi complessi. Ciò è dovuto all'uso diffuso del meccanismo del calco nei processi neologici di creazione degli equivalenti nelle diverse lingue ufficiali dell'Unione europea.

Questo analogo comportamento linguistico potrebbe indubbiamente facilitare lo sviluppo di soluzioni per il trattamento automatico del linguaggio. Per questo è necessario definire criteri di composizione neologica per standardizzare il lessico istituzionale, considerando, però, come abbiamo osservato, le questioni di adattamento alle specifiche tipologie di formazione lessicale delle lingue ufficiali dei differenti Stati membri e/o degli Stati che si relazionano con l'Ue. Pertanto, la creazione di un vocabolario controllato dei termini istituzionali di dominio per ogni lingua ufficiale dell'Unione europea con le loro

varianti lessicali sarebbe utile ai processi d'indicizzazione, di traduzione assistita e allo sviluppo di applicativi per il web semantico.

Come rileva il nostro studio, sia nei testi legali che in quelli informativi sono presenti termini specialistici del dominio osservato che veicolano concetti giuridicamente rilevanti, il cui significato è presente nell'apposita sezione "Definizioni" nei testi dei Regolamenti dell'Unione europea. Pertanto, questa attestazione ufficiale fa sì che questi termini rappresentino un lessico istituzionale che facilmente potrebbe diventare un vocabolario controllato prodotto dalle stesse istituzioni ed organismi dell'Unione europea.

Il passo successivo del nostro studio è trasformare il repertorio dei termini di questa prima indagine in una risorsa linguistica formato *Linguistic Linked Open Data*, in lingua italiana e nelle due varietà nazionali del portoghese (PE e PB), utile per lo sviluppo di applicativi per il web semantico.

Bionota:

Daniela Vellutino è ricercatrice di linguistica generale e glottologia nell'Università di Salerno, dove insegna Comunicazione pubblica e linguaggi istituzionali e Linguistica applicata alla comunicazione delle organizzazioni. È responsabile dell'Osservatorio sulla Comunicazione nella Pubblica Amministrazione in Italia e in Europa. È componente del consiglio scientifico dell'Associazione Italiana per la Terminologia. Partecipa alle attività della Rete per l'Eccellenza dell'Italiano istituzionale e alla Rete panlatina di terminologia (Realiter). È attivista nei movimenti civici per l'Open Government. Ha curato il volume "Comunicare in Europa. Lessici istituzionali e terminologie specialistiche" (Educatt, 2015).

Gian Luigi De Rosa è professore associato di Lingua e Traduzione – Lingue Portoghese e Brasiliana presso l'Università del Salento. Presidente della V edizione del SIMELP - SIMPÓSIO MUNDIAL DE ESTUDOS DE LÍNGUA PORTUGUESA (tenutosi a Lecce nel 2015, <http://www.simelp.it/>). Dal 2015 è direttore della Cattedra I. Camões-Unisalento "Manoel de Oliveira" e dal 2014 è co-direttore e co-responsabile scientifico della Unisalento Summer School of Audiovisual Translation. È autore di diversi saggi dedicati alla lingua, alla linguistica portoghese e alla traduzione audiovisiva e intersemiotica ed è traduttore letterario e audiovisivo.

Indirizzi e-mail: dvelluti@unisa.it; gianluigi.derosa@unisalento.it

Riferimenti bibliografici

- Berners-Lee T., *Linked Data: Principles and State of the Art* <https://www.w3.org/DesignIssues/LinkedData.html> (30.07.2017).
- Bombi R. 2005, *La linguistica del contatto. Tipologie di anglicismi nell'italiano contemporaneo e riflessi metalinguistici*, il Calamo, Roma.
- Church K.W. e Hanks P. 1990, *Word association norms, mutual information, and lexicography*, in "Computational Linguistics" 16 [1], pp. 22-29.
- De Mauro T. 1999, *Grande Dizionario italiano dell'uso*, UTET, Torino.
- De Rosa G.L. 2008, *Unimultiplicidade e Policentrismo do Português no Século XXI*, in Akta Konferencji "Dialogi z Luzofonią", Instytut Studiów Iberyjskich i Iberoamerykańskich UW, Warszawa, pp. 342-357. <http://iberystyka-uw.home.pl/content/view/345/113/lang.en/>.
- De Rosa G.L. 2009, *Interagire negli spazi discorsivi lusofoni: l'uso del VOCÊ nel portoghese europeo e nel portoghese brasiliano*, in De Rosa G. e De Laurentiis A. (a cura di), *Lingue Policentriche a contatto: quando la periferia diventa centro*, Polimetrica Publisher, Monza, pp. 73-88.
- Dunning T. 1993, *Accurate methods for the statistics of surprise and coincidence*, in "Computational Linguistics" 19 [1], pp. 61-74.
- Elia A. 1995, *Per una disambiguazione semi-automatica di sintagmi composti: i dizionari elettronici lessico-grammaticali*, in Cipriani R. e Bolasco S. (a cura di), *Ricerca qualitativa e computer*, FrancoAngeli, Milano, pp. 112-141.
- Frantzi K.T. e S. Ananiadou 1999, *The C-value/NC-value domain-independent method for multi-word term extraction*, in "Journal of Natural Language Processing" 6 [3], pp. 145-179.
- Gardner J. 2013, *A brief list of misused English terminology in EU publications*, European Court of Auditors.
- Sag I.A., Baldwin T., Bond F., Copestake A. and Flickinger D. 2002, *Multiword Expressions: A Pain in the Neck for NLP*, in "Lecture notes in Computer Science" 2276, pp. 1-15.
- Sager J.C. 1990, *A Practical Course in Terminology Processing*, John Benjamins, Amsterdam/Philadelphia.
- Salton G. e Buckley C. 1988, *Term-Weighting Approaches in Automatic Text Retrieval*, in "Processing & Management: an International Journal" 24 [5], pp. 513-523.
- Vellutino D., Elia A. e Marano F. 2012, *L'italiano istituzionale e le sue varietà d'uso pubblico. Aspetti lessicali nei tipi di testo d'informazione e comunicazione delle pubbliche amministrazioni*, in Bianchi P., De Blasi N., De Caprio C. e Montuori F. (a cura di), *La variazione nell'Italiano e nella sua storia. Varietà e varianti linguistiche e testuali*, Franco Cesati, Firenze, pp. 539-550.
- Vellutino D. 2014, *Esercizi di stile per il diritto di accesso civico*, in Macaluso F.P. (a cura di), *La lingua variabile nei testi letterari, artistici e funzionali contemporanei. Analisi, interpretazione traduzione*, Centro di Studi Filologici e Linguistici Siciliani, Palermo, pp. 172-174.
- Vellutino D. e Zanola M.T. 2015, *Comunicare in Europa. Lessici istituzionali e terminologie specialistiche*, in Vellutino D. e Zanola M.T. (a cura di), *Comunicare in Europa. Lessici istituzionali e terminologie specialistiche*, Educatt, Milano, pp. 7-10.
- Voghera M. 2004, *Polirematiche*, in Grossmann M. e Rainer F. (a cura di), *La formazione delle parole in italiano*, Max Niemeyer Verlag, Tübingen, pp. 56-69.