

Alma Mater Studiorum Università di Bologna

SCUOLA DI LINGUE E LETTERATURE, TRADUZIONE E INTERPRETAZIONE  
Sede di Forlì

Corso di Laurea Magistrale in Traduzione Specializzata (classe LM - 94)

TESI DI LAUREA

in Terminology and Information Mining

***Teoria Senso-Testo e Funzioni Lessicali.  
Una proposta per il trattamento delle collocazioni linguistiche***

CANDIDATO:  
Isabella Carini

RELATORE:  
Adriano Ferraresi

CORRELATORE:  
Silvia Bernardini

CORRELATORE ESTERNO:  
François Lareau  
(OLST - Université de Montréal)

Anno Accademico 2016/2017

Terza sessione



Le bocche muoiono  
La lingua è viva

(Bruno Tognolini, *Rima per gli ottant'anni di Tullio de Mauro*)





2.4.2.1	Funzioni lessicali standard vs. non standard .....	70
2.4.2.2	Funzioni lessicali sintagmatiche vs. paradigmatiche.....	71
2.4.2.3	Funzioni lessicali semplici e complesse; configurazioni di funzioni lessicali .....	73
2.5	Approfondimenti su alcune funzioni lessicali .....	76
2.5.1	Oper, Func, Labor e i verbi di supporto.....	77
2.5.1.1	Oper .....	77
2.5.1.2	Func .....	80
2.5.1.3	Labor.....	82
2.5.2	Real, Fact, Labreal e i verbi di realizzazione .....	84
2.5.2.1	Real.....	84
3.	Corpora e annotazione di corpora .....	89
3.1	Struttura del capitolo .....	89
3.2	I corpora.....	89
3.2.1	Definizione.....	89
3.2.2	Linguistica dei corpora.....	89
3.2.3	Criteri di valutazione dei corpora .....	91
3.2.4	Parametri di classificazione e tipi di corpora .....	93
3.2.5	Strumenti per la consultazione e l'analisi di corpora .....	96
3.3	Annotazione di corpora.....	97
3.3.1	Definizione e caratteristiche generali .....	97
3.3.2	Scopo dell'annotazione .....	100
3.3.3	Livelli di annotazione .....	100
3.3.3.1	Annotazione per parti del discorso .....	100
3.3.3.2	Lemmatizzazione .....	101
3.3.3.3	Annotazione sintattica .....	102
3.3.3.4	Annotazione semantica .....	103
3.3.3.5	Altri tipi di annotazione .....	104
4.	Studio di caso: uso delle funzioni lessicali nell'annotazione di corpora .....	107
4.1	Struttura del capitolo .....	107
4.2	Introduzione allo studio di caso .....	107
4.2.1	Le collocazioni nell'apprendimento linguistico .....	108
4.2.2	Risorse per l'apprendimento delle collocazioni: i dizionari specializzati .....	109
4.2.3	Risorse per l'apprendimento delle collocazioni: i corpora.....	111
4.2.4	Conclusioni sulle risorse passate in rassegna .....	111
4.2.5	La proposta oggetto del presente studio di caso.....	112
4.3	Il corpus EPTIC.....	113
4.3.1	Caratteristiche e struttura di EPTIC.....	114
4.3.2	Selezione del materiale linguistico da annotare .....	116
4.4	Annotazione per funzioni lessicali.....	119

4.4.1	Schema di annotazione .....	120
4.4.1.1	Tagset dell'annotazione.....	121
4.4.2	Schema di codifica.....	128
4.4.3	Modalità di consultazione dell'annotazione .....	130
4.5	Conclusioni sullo studio di caso .....	131
	Conclusioni.....	137
	Bibliografia .....	139
	Appendice 1: Glossario.....	I
	Appendice 2: Funzioni Lessicali Standard Semplici .....	XI
	Appendice 3: Scheda del lessema DIRITTO <sub>2</sub> .....	XXVII



## Abbreviazioni e convenzioni grafiche

DEC	Dizionario esplicativo-combinatorio
FL	Funzione lessicale
RSem	Rappresentazione semantica
RSintP	Rappresentazione sintattica profonda
RSintS	Rappresentazione semantica di superficie
RMorfP	Rappresentazione morfologica profonda
RMorfS	Rappresentazione morfologica di superficie
RFonP	Rappresentazione fonologica profonda
RFonS	Rappresentazione fonologica di superficie
SSem	Struttura semantica
SSem-com	Struttura semantico-comunicativa
SSintP	Struttura sintattica profonda
↔	corrispondenza (tra livelli di rappresentazione linguistica adiacenti)
≡	equivalenza (tra rappresentazioni appartenenti allo stesso livello)
'diritto'	un'espressione in minuscolo, preceduta e seguita da apici, rappresenta nella letteratura Senso-Testo (e nel Dizionario esplicativo-combinatorio) un semantema, oppure indica il senso di una parola o di un'espressione.
'A + B'	anche nella notazione formale di tipo matematico usata nella letteratura Senso-Testo gli apici indicano il senso. In questo esempio, la notazione corrisponde a "il senso di A + B".
DIRITTO	una parola in maiuscolo si riferisce nella letteratura Senso-Testo ad un LESSEMA (in particolare nei grafi della rappresentazione sintattica).
*dormire su molti allori	l'asterisco è la convenzione grafica utilizzata per indicare forme erranee di parole o espressioni.

***attante***<sup>α</sup>

il grassetto corsivo accompagnato dal simbolo  $\alpha$  in apice è la convenzione grafica utilizzata per indicare i termini ripresi nel glossario in Appendice 1.

## Résumé

L'objectif général de cette étude théorique est de présenter la théorie Sens-Texte, peu connue en Italie : la littérature en langue italienne sur ce sujet est très limitée et elle date de plus de vingt ans. Nous offrons une nouvelle introduction à cette théorie en italien, tout en proposant une terminologie raisonnée et uniforme. Un autre objectif important est de présenter le formalisme des fonctions lexicales et de souligner son utilité dans l'étude des collocations linguistiques. Les collocations jouent un rôle clé dans l'apprentissage des langues : leur maîtrise est, d'une part, un élément essentiel d'une compétence linguistique élevée, mais d'autre part cela pose des difficultés même aux étudiants de niveau avancé. L'étude de cas contenue dans ce travail vise à montrer comment les fonctions lexicales peuvent faciliter l'apprentissage des collocations, et cela à travers une proposition méthodologique pour l'annotation des collocations en corpus. L'étude de cas consiste en une annotation par fonctions lexicales des sous-corpus italiens de EPTIC (*European Parliament Translation and Interpreting Corpus*). Nous proposons une méthode pour étiqueter certaines collocations selon les catégories sémantiques fournies par les fonctions lexicales. Le corpus annoté et enrichi d'information sémantique deviendrait ainsi une ressource utile dans l'apprentissage des collocations de la langue italienne, en particulier pour les traducteurs et interprètes non italophones. Cette étude s'adresse aux linguistes et à tous les sujets qui souhaitent se familiariser avec la théorie Sens-Texte, aux lexicographes, aux professeurs de langues et à tous étudiants et chercheurs intéressés par une approche non conventionnelle des collocations.



## Resumen

Este estudio teórico tiene como objetivo general presentar la teoría lingüística Sentido-Texto de Igor' Mel'čuk, una teoría poco conocida en Italia, donde la literatura sobre este tema es limitada y data de hace más de veinte años. Por esta razón se quiere ofrecer una introducción actualizada a la teoría Sentido-Texto en italiano, y proponer una terminología uniforme y razonada. Otro objetivo importante es presentar el formalismo de las funciones léxicas del modelo Sentido-Texto y destacar su utilidad para el estudio de las colocaciones lingüísticas. Las colocaciones desempeñan un papel clave en el aprendizaje de idiomas, ya que su uso correcto es, por un lado, un factor esencial para lograr una alta competencia lingüística, pero por otro lado es un elemento de dificultad incluso para los estudiantes de nivel avanzado. Este estudio tiene como objetivo mostrar cómo las funciones léxicas se pueden utilizar para facilitar el aprendizaje de las colocaciones. En particular, el estudio presenta una propuesta metodológica para el tratamiento de las colocaciones en un corpus: una anotación de los subcorpus italianos del corpus EPTIC (*European Parliament Translation and Interpreting Corpus*), que consiste en etiquetar las colocaciones según las categorías semánticas proporcionadas por las funciones léxicas. De esta forma el corpus, enriquecido con información semántica, constituiría una herramienta útil para aprender las colocaciones del italiano, especialmente para estudiantes de traducción e interpretación que no son nativos de italiano.

Este estudio se dirige a lingüistas, y más en general a quién desee acercarse a la teoría Sentido-Texto, a los lexicógrafos, a los profesores de idiomas y a otros académicos interesados en un enfoque no convencional al estudio de las colocaciones.



## Abstract

The main purpose of this theoretical study is to introduce Igor' Mel'čuk's Meaning-Text Theory, which is little known in Italy, as the Italian literature on this subject is very limited and more than twenty years old. For this reason, we would like to provide an updated introduction to the Meaning-Text Theory in Italian, featuring a reasoned, uniform Italian terminology. Another important aim is to introduce lexical functions, a tool developed within the Meaning-Text framework, and to emphasize how they can be useful in the study of linguistic collocations. Collocations play a key role in language learning: on the one hand, mastering collocations in a foreign language is essential to achieve a high level of proficiency, but on the other hand collocations can be a stumbling block even for advanced learners.

The case study contained in this work shows a new method to add semantic information regarding collocations to a corpus: an annotation of the Italian subcorpora of EPTIC (*European Parliament Translation and Interpreting Corpus*), that is to say a method to label collocations according to the semantic categories provided by Mel'čuk's lexical functions. The annotated corpus, enriched with semantic information, would thus become a useful resource to learn Italian collocations, especially suited for non-Italian-speaking translators and interpreters. This study is addressed to linguists and more generally to anybody who wants to have a synthetic overview of the Meaning-Text theory, to lexicographers, to language teachers and to other scholars and students who might be interested in a non-mainstream approach to the study of collocations.



# Introduzione

Il presente studio teorico ha lo scopo generale di presentare la teoria linguistica Senso-Testo. In particolare si intende offrire un'introduzione aggiornata alla teoria Senso-Testo in lingua italiana, destinata idealmente a quanti non conoscono questa teoria e vi si vogliono avvicinare. La teoria Senso-Testo è infatti una teoria linguistica *non mainstream* ed è poco conosciuta in Italia, dove la letteratura a riguardo è limitata e, soprattutto, datata. Proprio per questo motivo si è deciso di dedicare una parte consistente di questo elaborato alla presentazione della teoria nei suoi diversi aspetti. Inoltre, poiché la stesura del presente elaborato ha richiesto un paziente lavoro di adattamento e uniformazione della terminologia Senso-Testo in italiano, la presentazione di una terminologia uniforme costituisce uno scopo ulteriore del presente lavoro. Un altro importante scopo è presentare il formalismo delle funzioni lessicali del modello Senso-Testo e sostenerne l'utilità per lo studio e il trattamento delle collocazioni linguistiche dell'italiano. Una collocazione linguistica, secondo la definizione proposta dalla teoria Senso-Testo, è un'espressione semi-idiomatica formata da due o più parole (la base e i suoi collocati) che è frutto di una composizione in cui la scelta della base determina la scelta dei collocati in virtù di restrizioni di tipo lessicale.

L'idea alla base del presente lavoro nasce da un'esperienza di studio presso il dipartimento di Linguistica e Traduzione dell'Université de Montréal (Canada), dove la teoria Senso-Testo è oggetto di ricerca e sviluppo da oltre trent'anni.

La teoria Senso-Testo è stata scelta come riferimento del presente lavoro per diverse ragioni. Innanzitutto perché essa offre un modello teorico completo del funzionamento delle lingue naturali, includendo tutti i livelli di analisi: semantica, sintassi, morfologia, fonologia e fonetica. Inoltre, la teoria Senso-Testo ha un approccio estremamente analitico allo studio della fraseologia e delle collocazioni linguistiche: il formalismo delle funzioni lessicali offre una descrizione sistematica e altamente formalizzata del fenomeno delle collocazioni. È soprattutto quest'ultimo aspetto ad aver stimolato l'interesse di chi scrive.

Il capitolo 1 è dedicato alla presentazione della teoria Senso-Testo nel suo complesso. Dopo un inquadramento storico-teorico generale si affrontano i caratteri della teoria Senso-Testo, il modello formale che da essa deriva e infine una breve trattazione sul lessico e la sua formalizzazione all'interno della teoria Senso-Testo.

Il capitolo 2 introduce il concetto di collocazione linguistica e presenta nel dettaglio le funzioni lessicali, lo strumento offerto della teoria Senso-Testo per descrivere e rappresentare formalmente le collocazioni.

Il capitolo 3 presenta da un punto di vista teorico l'oggetto dello studio di caso descritto nel capitolo 4, ossia l'annotazione di corpora: la prima parte del capitolo è dedicata all'introduzione dei corpora e dei loro possibili utilizzi per scopi di studio e ricerca; la seconda parte del capitolo definisce natura e scopi dell'annotazione di corpora.

Il capitolo 4 costituisce infine uno studio di caso il cui scopo è proporre un'applicazione pratica delle funzioni lessicali nel trattamento delle collocazioni. Lo studio di caso consiste nella realizzazione di un'annotazione per funzioni lessicali dei sottocorpora italiani del corpus EPTIC (*European Parliament Translation and Interpreting Corpus*), ossia un modo per etichettare alcune collocazioni in base alle categorie semantiche fornite dalle funzioni lessicali. In questo modo il corpus, così arricchito di informazioni, costituirebbe uno strumento utile per l'apprendimento di alcune collocazioni della lingua italiana, destinato in particolare agli studenti di traduzione e interpretazione non italofofi.

I materiali in appendice costituiscono una parte importante del presente lavoro. Essi sono stati concepiti e realizzati sia come supporto alla lettura dei capitoli, sia come strumento per la consultazione indipendente, destinato a chi voglia approcciarsi alla teoria Senso-Testo e desidera avere un inquadramento sintetico ma completo dei principali concetti. In particolare il glossario in Appendice 1 riprende e approfondisce i concetti di base della teoria Senso-Testo trattati nei vari capitoli, mentre l'Appendice 2 è una tabella in cui sono descritte in sintesi tutte le funzioni lessicali individuate da Mel'čuk, anche quelle che per ragioni di spazio non sono state menzionate nel capitolo 2. L'Appendice 3, infine, costituisce una bozza di scheda lessicografica redatta secondo i principi della lessicografia Senso-Testo, la cui realizzazione è stata necessaria per poter condurre lo studio di caso del capitolo 4.

# 1. La teoria Senso-Testo

## 1.1 Struttura del capitolo

Il presente capitolo è strutturato come segue: dopo una necessaria premessa terminologica (sezione 1.2) sono illustrati il contesto storico della teoria Senso-Testo e i suoi sviluppi fino ai giorni nostri (sezione 1.3); nella sezione 1.4 sono presentati i caratteri e i postulati della teoria Senso-Testo; la sezione 1.5 presenta invece il modello Senso-Testo e due dei suoi livelli di rappresentazione. Il capitolo si chiude con una breve presentazione del Dizionario esplicativo-combinatorio, ossia la componente lessicale del modello Senso-Testo (sezione 1.6).

## 1.2 Premessa terminologica

La teoria Senso-Testo fa uso di una terminologia propria, che comprende sia neologismi sia termini già noti nel dominio della linguistica (che assumono, in questo contesto, significati diversi). La dimensione terminologica ha dunque un'importanza non secondaria nella presentazione teorica.

Nelle pagine che seguono saranno presentati i fondamenti della teoria Senso-Testo, e contestualmente saranno offerte le definizioni dei concetti principali, riprese e approfondite in un glossario riportato in Appendice 1.

I termini contenuti nel glossario sono segnalati in **grassetto corsivo** e seguiti dal simbolo  $\alpha$  in apice. Esempio: **modello funzionale $\alpha$** .

È necessario precisare, inoltre, che la terminologia italiana *attestata* relativa alla linguistica Senso-Testo è molto limitata. Il motivo è che la maggior parte della letteratura Senso-Testo è in lingua inglese, francese e russa. Gli unici testi in lingua italiana che descrivono l'impianto della teoria Senso-Testo sono Rigotti (1979), Gatti (1992) e Prencipe (2006). Questi testi costituiscono i miei riferimenti terminologici principali, ma non sono sufficienti a coprire tutti i concetti illustrati nel presente lavoro; inoltre, alcuni termini italiani già attestati in Gatti (1992) si sono rivelati poco efficaci perché riferiti alla prima formulazione della teoria Senso-Testo (Mel'čuk, 1974), ormai datata. Per questi motivi è stato necessario sopperire alle lacune introducendo nuovi termini in italiano.

In sintesi, la terminologia del presente lavoro è stata stabilita secondo i seguenti criteri:

- si è cercato di limitare al massimo il ricorso ad anglicismi, utilizzandoli solo in forma di citazione o nelle note terminologiche esplicative;
- ove possibile, sono stati impiegati termini (e relative abbreviazioni) già attestati in Rigotti (1979) e Gatti (1992);
- in caso di mancanza o inadeguatezza di un termine italiano attestato, il termine è stato coniato traducendolo dall'inglese o dal francese (le due lingue in cui sono scritti i testi più recenti e autorevoli della letteratura Senso-Testo);
- nei casi in cui tra la versione inglese e francese dello stesso termine ci sia una grossa differenza, per coniare il termine italiano si è preferito mantenersi aderenti alla morfologia del francese, per una questione di uniformità.

### 1.3 Contesto storico

L'impianto teorico su cui si basa il presente lavoro è la teoria<sup>1</sup> linguistica Senso-Testo, elaborata a partire dagli anni sessanta da vari studiosi russi tra cui Aleksandr Konstantinovič Žolkovskij, Jurij Derenikovič Apresjan e in particolare Igor' Aleksandrovič Mel'čuk, dal cui contributo la teoria dipende in gran parte, tanto da essere legata principalmente al suo nome.

L'originalità dell'approccio di Mel'čuk risiede in una molteplicità di fattori, tra cui il ruolo centrale assunto dalla semantica, l'attenzione riservata al lessico e alla sua formalizzazione, la capacità di integrare spunti provenienti dai maggiori indirizzi della linguistica del ventesimo secolo e, non ultime, le possibili applicazioni computazionali del modello Senso-Testo, ossia il modello formale che è il prodotto della teoria Senso-Testo.

---

<sup>1</sup> Benché il lavoro di Mel'čuk e Žolkovskij sia generalmente conosciuto con il nome di *Meaning-Text theory*, Mel'čuk stesso ha specificato che non si tratta di una *teoria* in senso stretto (come insieme di teoremi verificabili) ma che il termine sia piuttosto da intendersi nell'accezione che esso ha nelle scienze umane. Mel'čuk suggerisce due alternative terminologiche: *Meaning-Text approach* per l'impianto teorico generale (a cui comunemente ci si riferisce con *teoria*) e *Meaning-Text model* per il modello formale che da esso deriva (cfr. Mel'čuk, 2012: 36). Dal momento che nella letteratura è ampiamente attestato il termine *Teoria Senso-Testo*, nel presente lavoro si farà preferibilmente uso di questa denominazione, mentre il termine *modello Senso-Testo* sarà utilizzato esclusivamente per riferirsi al modello formale che da tale teoria deriva.

### 1.3.1 Gli sviluppi della linguistica sovietica negli anni sessanta

Gli anni sessanta del ventesimo secolo rappresentarono un momento di enorme sviluppo per la linguistica di area sovietica, dovuto a un concorso di circostanze: un allentamento della censura che favorì l'apertura alla linguistica occidentale, in particolare allo strutturalismo,<sup>2</sup> e l'interesse crescente verso gli approcci matematici e formali in linguistica (Versace, 2011). Fu precisamente in tale periodo che emerse in URSS il filone teorico definito *linguistica matematica*, nato da diverse istanze: a livello epistemologico, dalla volontà di dare alla linguistica il rigore e la dignità scientifica propria delle altre scienze (attraverso l'introduzione di metodi quantitativi); a livello pratico, dall'urgenza di sfruttarne le applicazioni nel campo nascente della traduzione automatica e nell'automazione dei processi informativi (Gatti, 1992). A partire dal 1956 nell'ambiente accademico moscovita si moltiplicarono le iniziative volte alla valorizzazione delle prospettive matematiche in linguistica, tra cui ricordiamo, assieme a conferenze e seminari specifici, la nascita dell'Associazione per la traduzione meccanica, diretta da V. Rozencveijg presso l'Istituto di Lingue Straniere di Mosca (1956), e la fondazione dell'Istituto di Linguistica Matematica presso l'Università di Mosca (1960).

L'interesse per le applicazioni pratiche, fortemente incoraggiato dalle autorità, porterà la linguistica matematica sovietica a svilupparsi in molteplici direzioni, raggruppabili entro due linee fondamentali: da una parte l'analisi formale delle lingue naturali, e dall'altra l'ideazione di lingue artificiali al servizio dei primi esperimenti di traduzione automatica (concetto di *lingua mediatrice* o *interlingua*) (Gatti, 1992). È possibile e doveroso operare un'ulteriore distinzione interna all'analisi formale delle lingue naturali, che corrisponde a due approcci distinti: un approccio statistico, basato su evidenze estratte dall'analisi di corpora,<sup>3</sup> e un approccio strutturale, definito anche *algebrico* (Gatti, 1992), che studia e descrive il sistema linguistico

---

<sup>2</sup> A partire dagli anni venti i metodi e contenuti dell'indagine linguistica in URSS avevano subito pesanti condizionamenti di natura ideologica; in particolare gli indirizzi di ricerca orientati all'approccio di Saussure furono dichiarati incompatibili con l'ideale marxista e abbandonati. Solo nel 1956 lo strutturalismo tornò a essere oggetto di dibattito, e ne fu infine riconosciuto il valore, grazie alla pubblicazione di un saggio di S.K. Šaumjan sul numero V della rivista "Voprosy Jazykoznanija" (rivista di linguistica fondata per volontà di Josif Stalin nel 1952). Per una trattazione approfondita del dibattito sullo strutturalismo in URSS si rimanda a Rigotti (1973: 488-521). Questa apertura teorica non segnò tuttavia la fine del pesante controllo del regime comunista sulla scienza linguistica, come testimonia la vicenda accademica e umana dello stesso Mel'čuk, esposta in Gatti (1992: 7-8).

<sup>3</sup> Per una rassegna delle ricerche linguistiche di indirizzo statistico in Unione Sovietica si rimanda al capitolo 4 di Papp (1966).

facendo ricorso a strumenti matematici quali la logica e la teoria degli insiemi per la costruzione di modelli formali che descrivano il funzionamento della lingua. La teoria Senso-Testo di Mel'čuk è da collocarsi in quest'ultima tendenza.

Pur nella diversità di indirizzi, la linguistica sovietica degli anni sessanta si distinse per l'interesse verso "tematiche particolarmente importanti per la tradizione linguistica russa, come il rapporto tra lingua e pensiero e lo status della semantica all'interno delle discipline linguistiche" (Versace, 2011:5).

Benché, come affermato, il lavoro di Mel'čuk si inserisca sostanzialmente nel filone descrittivo delle lingue naturali, esso non è estraneo né agli aspetti applicativi legati alla traduzione automatica, né all'interesse per lo sviluppo di lingue mediatrici. Fin dagli anni cinquanta Igor' Mel'čuk si distinse tra i fondatori della traduzione automatica in URSS: nel 1954 diresse le ricerche per l'ideazione di un algoritmo traduttivo dal francese al russo presso l'Istituto di Matematica Steklov di Mosca; nel 1958 realizzò un algoritmo traduttivo ungherese-russo. Fu proprio nel corso di queste ricerche che Mel'čuk, assieme ai colleghi, ipotizzò la necessità di un sistema simbolico universale a cui poter ricondurre i sistemi espressivi delle varie lingue, un metalinguaggio che descrivesse l'organizzazione del senso sotteso alle realizzazioni linguistiche.

Gradualmente si fece strada l'idea che la lingua mediatrice deve corrispondere a una rappresentazione del senso comune a tutte le lingue, in altre parole, a una *lingua semantica universale*. Tale lingua semantica non era per altro intesa soltanto come strumento tecnico elaborato per la traduzione automatica, ma come aspetto reale della capacità linguistica del parlante (Rigotti 1979:295). In questa visione della lingua mediatrice si trovavano già i principi fondamentali di ciò che sarebbe stata la rappresentazione semantica del modello Senso-Testo.

### **1.3.2 Nascita della teoria Senso-Testo: dal Circolo Semantico di Mosca agli sviluppi contemporanei**

La Teoria Senso-Testo ebbe origine nei primi anni sessanta dalle intuizioni di un gruppo di ricercatori del Laboratorio per la traduzione meccanica dell'Istituto di Lingue Straniere di Mosca, fra cui Žolkovskij e Mel'čuk. In due articoli (Žolkovskij e Mel'čuk, 1965, 1967) essi segnalavano i limiti dell'approccio tradizionale alla traduzione automatica (basato sull'analisi

sintattica dei testi per la successiva trasposizione in lingua di arrivo), in particolare l'insufficiente gestione dei casi di omonimia, e proponevano un nuovo metodo, basato sulla formalizzazione semantica anziché sintattica, e su un algoritmo di sintesi che a partire da una rappresentazione del significato del testo fosse in grado di fornirne la traduzione. In questo modo, la traduzione automatica avrebbe riprodotto l'attività del traduttore umano, che non traduce parole o strutture, bensì estrae il significato delle espressioni per trasporle nella lingua di arrivo. In questa proposta metodologica erano contenuti *in nuce* i concetti che in seguito avrebbero costituito l'impianto teorico Senso-Testo. Nella formulazione di questa proposta fu grande il contributo di J. Apresjan, studioso di semantica già affermato all'epoca, che nel 1967 si unì ai lavori di Žol'kovskij e Mel'čuk dando vita al nucleo del cosiddetto *Circolo Semantico di Mosca*, un gruppo di linguisti che tra il 1967 e il 1977 lavorò alla costruzione del modello linguistico Senso-Testo. Il primo esito fu la pubblicazione, nel 1974, della monografia di Mel'čuk *Opyt teorii lingvističeskich modelej «Smysl-Tekst». Semantika, sintaksis* [Per una teoria dei modelli linguistici «Senso-Testo». Semantica, sintassi], una prima formulazione organica del modello. Nel corso degli anni 70, numerosi studiosi contribuirono allo sviluppo del modello Senso-Testo, e la ricerca fu orientata principalmente verso l'ambito della semantica e della lessicografia (Mel'čuk, 2013).

Il periodo di elaborazione del modello Senso-Testo coincise con quella che gli storici hanno definito "svolta antistrutturalista" o "*sudden reideologization*" in Unione sovietica (Gatti, 1992: 20), ossia un periodo di inasprimento ideologico che in ambito accademico portò ad ostacolare le ricerche di Mel'čuk, sia per il suo costante riferimento alla linguistica occidentale (in particolare allo Strutturalismo),<sup>4</sup> sia per le posizioni politiche assunte da Mel'čuk come individuo. Alla luce di questa premessa è possibile comprendere sia la mancata pubblicazione in URSS delle opere successive di Mel'čuk, sia la decisione di Mel'čuk, Žol'kovskij e altri di emigrare, dal 1977 in poi.

Tra la fine degli anni settanta e gli anni ottanta la teoria Senso-Testo iniziò a diffondersi in occidente, sia grazie alle ricerche dei teorici russi emigrati (svolte presso università europee ed americane), sia grazie alla pubblicazione di alcuni saggi e monografie da parte di editori

---

<sup>4</sup> Lo strutturalismo in URSS era già stato oggetto di opposizioni ideologiche nei decenni precedenti. Si veda nota 2 del presente capitolo.

occidentali: la casa editrice tedesca Fink Verlag nel 1976 pubblicò una raccolta di scritti di Mel'čuk sugli aspetti morfologici del modello Senso-Testo, la cui pubblicazione in URSS non era stata possibile (Mel'čuk, 1976); la stessa casa editrice nel 1982 pubblicò un altro volume fondamentale<sup>5</sup> per il modello, in cui Mel'čuk, "mosso dall'urgenza di superare l'incertezza terminologica, cui spesso si lega un'imprecisione concettuale" (Gatti, 1992:24), presentava un sistema concettuale e terminologico unitario e coerente per la parte morfologica del modello Senso-Testo.

Nel 1984 la rivista viennese di slavistica *Wiener Slawistischer Almanach* (già fautrice della pubblicazione di numerosi scritti del Circolo Semantico di Mosca) pubblicò il primo volume del Dizionario esplicativo-combinatorio<sup>6</sup> della lingua russa, frutto del lavoro pluridecennale del Circolo, uno dei contributi più interessanti e innovativi della teoria Senso-Testo agli studi linguistici e lessicografici (cfr. sezione 1.6 sul Dizionario esplicativo-combinatorio).

Gli anni ottanta rappresentarono inoltre un momento importante per la ricezione critica della teoria Senso-Testo in occidente: le prime pubblicazioni a carattere critico risalgono a questo decennio (Gatti, 1992).

Tra la metà degli anni ottanta e la fine degli anni novanta linguisti e studiosi guidati da Mel'čuk elaborarono il Dizionario esplicativo-combinatorio della lingua francese, risultato di ricerche e lavori iniziati presso l'Université de Montréal sin dal 1979 (Mel'čuk e al., 1984, 1988, 1992, 1999).

Negli anni novanta l'approccio Senso-Testo venne applicato in particolare allo studio della fraseologia: citiamo a tal proposito i lavori di Mel'čuk (1996, 1998), Cowie (1998), Wanner (1996), Alonso Ramos (1998). In questo periodo molti altri studiosi iniziarono a interessarsi all'approccio Senso-Testo, specialmente in vista delle possibili applicazioni del modello nel trattamento automatico delle lingue. Nel 1997 presso l'Université de Montréal venne creato

---

<sup>5</sup> I. A. Mel'čuk (1982) *Towards a language of linguistics. A system of formal notations for Theoretical Morphology*. München: Fink.

<sup>6</sup> *Dizionario esplicativo-combinatorio* è il termine adottato nel presente lavoro (e abbreviato in DEC). In italiano esisteva già un termine attestato: Gatti (1992) si riferisce all'opera di Mel'čuk con il termine *Vocabolario interpretativo-combinatorio*. Tuttavia, per ragioni di omogeneizzazione terminologica, nel presente lavoro si è preferito optare per una traduzione aderente al termine inglese *Explanatory Combinatorial Dictionary* (Mel'čuk, 2013) e al francese *Dictionnaire Explicatif et Combinatoire*.

l'Observatoire Linguistique Sens-Texte (OLST),<sup>7</sup> importante centro di ricerca in cui l'interesse per la modellizzazione formale dei fenomeni linguistici venne da subito applicato anche agli ambiti della didattica e della terminologia.

Nel corso degli anni duemila gli aspetti formali e le possibili applicazioni del modello Senso-Testo guadagnarono una sempre maggiore attenzione da parte della comunità accademica. Tra i contributi più rilevanti di questi anni ricordiamo, oltre a quelli dello stesso Mel'čuk, il lavoro svolto presso l'Observatoire Linguistique Sens-Texte da studiosi come Alain Polguère e Sylvain Kahane. Un importante progetto lessicografico portato a termine negli anni duemila è il *DiCo*, dizionario costruito sul modello del Dizionario esplicativo-combinatorio ma dedicato prevalentemente alla descrizione delle relazioni lessicali e alle co-occorrenze; il *DiCo* e il *DiCouèbe*,<sup>8</sup> la sua versione consultabile online, nacquero dal lavoro congiunto di ricercatori dell'Observatoire de Linguistique Sens-Texte dell'Université de Montréal e di studiosi francesi del laboratorio *Lattice* del CNRS (Jousse e Polguère, 2005). Nel 2007 fu pubblicato a nome di Mel'čuk e Polguère un nuovo dizionario per la lingua francese sul modello del Dizionario esplicativo-combinatorio, il *Lexique Actif du Français* (Mel'čuk e Polguère, 2007).

Dal 2010 a oggi si sono moltiplicati gli studi legati alla teoria Senso-Testo, e numerosi ricercatori hanno orientato i loro lavori verso questo campo di indagine. L'Observatoire de Linguistique Sens-Texte dell'Université de Montréal è tuttora uno dei centri più importanti, dove un nutrito gruppo di studiosi diretto da Patrick Drouin e François Lareau lavora sul modello Senso-Testo dal punto di vista sia teorico che pratico (applicazioni in lessicografia, in didattica, in terminologia, applicazioni computazionali, generazione automatica di testo e trattamento automatico dei testi).

#### 1.4 Caratteristiche della teoria Senso-Testo

La teoria linguistica Senso-Testo, come già sottolineato da Rigotti (1979) e Gatti (1992), ha come base una concezione della lingua naturale come strumento destinato alla

---

<sup>7</sup> <http://olst.ling.umontreal.ca/>.

<sup>8</sup> <http://idefix.ling.umontreal.ca/dicouebe/>.

comunicazione (è evidente il legame con il pensiero di Jakobson (1960) sulla funzione eminentemente comunicativa del linguaggio).

Il processo comunicativo è visto in questo contesto come un passaggio da un *senso* a un *testo* (inteso qui semplicemente come significante linguistico, l'espressione verbale del senso) e viceversa, ossia come un processo in cui il mittente, a partire da un senso che desidera esprimere, arriva a costruire un testo che il destinatario dovrà analizzare per risalire al senso in esso contenuto. Il mezzo che permette di stabilire la corrispondenza tra sensi e testi è la lingua naturale.

La teoria Senso-Testo è una teoria linguistica descrittiva e universale (che si applica cioè a tutte le lingue naturali), che fornisce gli strumenti per la costruzione di modelli funzionali delle lingue naturali (ossia modelli che rappresentano il funzionamento delle lingue), detti modelli Senso-Testo (Milićević, 2006). Diversamente da altre teorie linguistiche che descrivono il funzionamento della lingua principalmente attraverso la grammatica, relegando il lessico a un ruolo marginale, la teoria Senso-Testo attribuisce al lessico un ruolo fondamentale, come vedremo nella sezione 1.6.

Uno dei più interessanti elementi di novità introdotti dalla teoria Senso-Testo è la *terza articolazione del segno linguistico*<sup>α</sup>. Oltre al significato e al significante (intesi nell'accezione di Saussure, 1916), Mel'čuk individua un'ulteriore articolazione del segno linguistico, quella costituita dalle *proprietà combinatorie*<sup>α</sup> del segno,<sup>9</sup> ossia il modo in cui esso può combinarsi ad altri segni linguistici all'interno di una data lingua naturale (in altri termini, le co-occorrenze). Le proprietà combinatorie nella teoria Senso-Testo hanno un'importanza tale da essere ritenute non proprietà contingenti, bensì una caratteristica intrinseca e strutturale di un dato elemento linguistico, al pari di significante e significato.

---

<sup>9</sup> Nella letteratura Senso-Testo in lingua inglese la terza articolazione del segno linguistico (ossia le sue proprietà combinatorie) è indicata con il termine *syntactics* (da non confondere con *syntax*, =sintassi); nella letteratura in lingua francese il termine è il sostantivo maschile *syntactique* (da non confondere con l'aggettivo *syntaxique* = sintattico), oppure il termine *propriétés de combinatoire*. Non esiste un termine italiano attestato corrispondente all'inglese *syntactics*, ci serviremo quindi unicamente del termine *proprietà combinatorie*, attestato già in Gatti (1992), e morfologicamente simile al francese *propriétés de combinatoire*.

Il campo di indagine della teoria Senso-Testo è limitato agli aspetti strettamente linguistici della comunicazione, sono pertanto esclusi gli aspetti cognitivi e relativi alla facoltà del linguaggio in quanto tale (Mel'čuk, 2012).

#### 1.4.1 Postulati della teoria Senso-Testo

In questa sezione presenteremo i fondamenti della teoria Senso-Testo riferendoci alla presentazione di Mel'čuk (2012) e all'ottima sintesi fornita da Milićević (2006).

La teoria Senso-Testo si regge su tre postulati.

##### Postulato 1

Una lingua naturale è (considerata) un insieme finito di regole che specificano una corrispondenza “molti a molti” tra un insieme infinito numerabile di sensi e un insieme infinito numerabile di testi (Mel'čuk, 2012:110, traduzione mia)

Questo primo postulato indica l'oggetto centrale della teoria: la lingua naturale come corrispondenza tra sensi e testi.

I termini **sens**<sup>α</sup> e **testo**<sup>α</sup> sono da intendere, nel contesto della teoria, in un'accezione particolare: il senso è inteso in termini puramente linguistici (slegato da aspetti pragmatici, connotativi, culturali) come il contenuto di un'espressione che il mittente intende comunicare, e il testo è il significante linguistico, il segnale fisico (orale, scritto o gestuale) attraverso cui il senso è trasmesso. Senso e testo sono dunque i *dati linguistici* oggetto dell'analisi (Milićević, 2006).

La corrispondenza tra sensi e testi è molti a molti perché un senso può essere espresso da diversi testi (come nei casi di sinonimia), e uno stesso testo può avere diversi sensi (ambiguità, omonimia o polisemia).

È opportuno specificare che la teoria Senso-Testo non si occupa di sensi e testi reali, ma di *rappresentazioni* di sensi e testi, ossia di una

descrizione di essi attraverso linguaggi formali specificamente ideati a questo scopo. Un senso viene rappresentato mediante un oggetto formale chiamato *rappresentazione semantica* [=RSem] e, in maniera analoga, un testo viene rappresentato mediante una rappresentazione fonologica di superficie [= RFonS].

Dunque il postulato 1 può essere formulato mediante la seguente notazione simbolica: {RSem} <=lingua=> {RFonS}.<sup>10</sup> (Milićević 2006, traduzione mia)

Una rappresentazione dunque non è solo un metalinguaggio utilizzato per descrivere diversi livelli della lingua, ma è anche un oggetto formale dotato di caratteristiche precise, che saranno approfondite nella sezione 1.5.

Se il primo postulato indica l'oggetto della teoria (la lingua come corrispondenza tra sensi e testi), resta da capire come tale oggetto funziona. I linguisti si trovano qui di fronte al classico problema della “scatola nera” (Mel'čuk, 2012), poiché gli input e output della lingua, ossia sensi e testi, sono entità accessibili al linguista, ma il meccanismo che permette di passare dagli input agli output (ossia le regole di funzionamento della lingua, che in definitiva rendono possibile la comunicazione) resta sconosciuto. Questo problema viene affrontato dal secondo postulato.

#### **Postulato 2:**

Una lingua naturale L, vista come una corrispondenza tra sensi e testi, deve essere descritta da un modello formale particolare, detto modello funzionale (Mel'čuk, 2012, elaborazione mia).

La corrispondenza tra sensi e testi è descritta da un **modello funzionale**<sup>α</sup>, cioè un insieme di regole che simula il funzionamento di un fenomeno – in questo caso l'attività linguistica di un parlante nativo – a partire dai suoi input e output, cercando di ricostruire i processi e le regole alla base del suo funzionamento.

Il *modello funzionale Senso-Testo* è dunque un insieme di regole che simulano il processo attraverso cui, a partire da un senso, si arriva alla produzione di testi (*sintesi linguistica*) e il processo attraverso cui, viceversa, a partire da un testo si risale al senso in esso contenuto (*analisi linguistica*). È importante sottolineare l'aspetto di simulazione sopra accennato: il modello Senso-Testo, come ogni modello funzionale, *simula* il funzionamento della lingua, senza pretesa di avere validità da un punto di vista psicolinguistico: non esistono ancora dati

---

<sup>10</sup> Nel presente lavoro, oltre alla terminologia, anche la notazione formale è adattata alla lingua italiana, come già in Rigotti (1979) e Gatti (1992). Le abbreviazioni qui utilizzate per la notazione formale fanno dunque riferimento all'ordine testa-modificatore della lingua italiana (Rappresentazione Semantica = RSem, vs. Semantic Representation = SemR).

sufficienti in grado di dimostrare che tale modello riproduca effettivamente i processi cognitivi che hanno luogo nel cervello dei parlanti.

A questo punto il problema che si pone è legato alla complessità del fenomeno oggetto di studio. La rappresentazione semantica (livello del senso) e la rappresentazione fonologica (livello del testo) non bastano a rendere conto delle corrispondenze tra sensi e testi: occorre definire dei livelli di rappresentazione intermedi. Il terzo postulato fa luce proprio su questo aspetto.

### **Postulato 3**

per descrivere la corrispondenza  $\{RSem\} \Leftrightarrow \{RFon\}$  di una lingua naturale L, si introducono due livelli di rappresentazione intermedi:

- la rappresentazione sintattica [=RSint], che descrive i fenomeni operanti a livello di organizzazione della frase
- la rappresentazione morfologica [=RMorf], che descrive i fenomeni operanti a livello di parola. (Mel'čuk, 2012:122, elaborazione mia)

Tutti i livelli di rappresentazione, fatta eccezione per il livello semantico, sono distinti in due sotto-livelli: livello profondo e livello di superficie. Il modello Senso-Testo conta quindi sette livelli di rappresentazione linguistica:

- rappresentazione semantica [=RSem];
- rappresentazione sintattica profonda [=RSintP];
- rappresentazione sintattica di superficie [=RSintS];
- rappresentazione morfologica profonda [=RMorfP];
- rappresentazione morfologica di superficie [=RMorfS];
- rappresentazione fonologica profonda [=RFonP];
- rappresentazione fonologica di superficie [=RFonS].

Il postulato 3 chiarisce dunque che il passaggio dal Senso al Testo (attraverso la lingua naturale) non è diretto, bensì articolato in sette passaggi intermedi, corrispondenti ai sette livelli di rappresentazione linguistica del modello. In altre parole, per “tradurre” un senso in espressione verbale occorre compiere una serie di passaggi attraverso i vari livelli della lingua, dalla semantica, alla sintassi, alla morfologia, alla fonologia, dove ad ogni passaggio si compiono una serie di codifiche specifiche per quel livello. Tale serie di passaggi può essere

formalizzata e riassunta mediante la seguente notazione, che descrive anche, come vedremo, l'architettura di base del modello Senso-Testo:

$$\{\text{RSem}\} \Leftrightarrow \{\text{RSintP}\} \Leftrightarrow \{\text{RSintS}\} \Leftrightarrow \{\text{RMorfP}\} \Leftrightarrow \{\text{RMorfS}\} \Leftrightarrow \{\text{RFonP}\} \Leftrightarrow \{\text{RFonS}\}$$

Si noti che nella notazione proposta le frecce sono bidirezionali: la serie di passaggi in cui è articolato il funzionamento della lingua nel modello Senso-Testo può infatti essere compiuta in due direzioni, perché la lingua permette sia di esprimere sensi, cioè passare da un senso alla sua espressione verbale in forma di testo (meccanismo di sintesi linguistica), sia di comprendere il senso di enunciati, cioè passare da un testo al suo senso (meccanismo di analisi linguistica).

Oltre ai tre postulati di base, Mel'čuk (2012) segnala anche un principio metodologico importante: nonostante la teoria Senso-Testo consenta, attraverso il modello Senso-Testo, di rappresentare il funzionamento della lingua sia nella direzione della sintesi sia nella direzione dell'analisi (come visto nel paragrafo precedente), la prospettiva di studio privilegiata da Mel'čuk è quella della sintesi, per una serie di motivi tra cui citeremo solo quello più rilevante ai fini del presente lavoro:

alcuni fenomeni linguistici sono più facilmente analizzabili dal punto di vista della sintesi linguistica; l'importanza e la difficoltà dello studio delle co-occorrenze lessicali (ad esempio collocazioni come *do a favor*, *make a mistake*, *file a complaint*, ecc.) emergono solo se considerate nella prospettiva che va dal senso al testo (Milićević, 2006, traduzione mia).

La teoria Senso-Testo si propone dunque di rispondere più alla domanda “come viene espresso un dato senso nella lingua L?”, invece che “che cosa significa un dato enunciato in una lingua L?”; come indica Milićević, ciò ha come corollario il fatto che lo studio dei meccanismi di parafrasi (che danno vita a espressioni linguistiche diverse aventi lo stesso senso) occupi un ruolo di estrema importanza all'interno dell'approccio Senso-Testo. Per un approfondimento sul tema della parafrasi, che non verrà toccato nel presente lavoro, si rimanda a Mel'čuk (2012:67-106) e a Milićević (2003).

## 1.5 Il modello funzionale Senso-Testo

### 1.5.1 Caratteristiche generali del modello Senso-Testo

Dopo aver esposto le basi della teoria Senso-Testo, ossia i suoi postulati, e alcune definizioni fondamentali, passeremo ora all'introduzione delle caratteristiche del modello Senso-Testo, il prodotto di tale teoria. Come sottolineato da Milićević (2006), le caratteristiche fondamentali del modello Senso-Testo sono una diretta conseguenza dei tre postulati della teoria Senso-Testo:

- Dal postulato 1 emerge che il modello Senso-Testo non è un modello generativo. Lo scopo del modello non è di generare tutte le espressioni grammaticalmente corrette per esprimere un dato senso, bensì descrivere come si arrivi da un senso a tutti i possibili testi in grado di esprimere quel senso. Mel'čuk descrive il modello in termini di *equative model* o *translative model*, sottolineando appunto che il modello si limita a individuare e descrivere le *corrispondenze* tra sensi e testi. Anche Gatti sottolinea il carattere *traduttivo* e non generativo del modello Senso-Testo: "In quanto sistema di corrispondenze tra sensi e testi, la lingua è sostanzialmente un meccanismo traduttivo: arrivare ad un testo a partire da un senso dato e viceversa non vuol dire altro che effettuare una serie di traduzioni, passando attraverso i livelli linguistici intermedi in cui si articola il passaggio dai primi ai secondi e viceversa" (Gatti, 1992:32).
- dal postulato 2 emerge la natura formale del modello: i sensi, le espressioni linguistiche e le regole che permettono di arrivare dagli uni alle altre sono rappresentati mediante un linguaggio formale.
- dal postulato 3 emerge la natura stratificata del modello:<sup>11</sup> esso possiede sette livelli di rappresentazione (corrispondenti ai diversi livelli di organizzazione linguistica) e sei *moduli*, ossia insiemi di regole di corrispondenza tra un livello e i livelli adiacenti a esso (Figura 1).

---

<sup>11</sup> L'idea di un modello linguistico stratificato non è nuova: negli anni sessanta già la grammatica stratificazionale del linguista americano Sydney Lamb e il modello generativo dell'ungherese Petr Sgall avevano introdotto questo principio, seppure in forme diverse.

Nei paragrafi successivi il modello Senso-Testo sarà descritto più in dettaglio, a partire dai livelli di rappresentazione e dai moduli di cui esso si compone.

Il passaggio da un livello di rappresentazione  $n$  al livello adiacente  $n+1$  avviene attraverso una componente del modello Senso-Testo chiamato modulo. Un modulo è un insieme di regole di corrispondenza (regole che determinano cioè come si stabilisce una corrispondenza), che a partire da un input – la rappresentazione di livello  $n$  – è in grado di “costruire” come output la rappresentazione di livello  $n+1$  (o, più precisamente, tutte le rappresentazioni possibili, tra loro equivalenti perché sinonimiche, di livello  $n+1$ ). In questo modo, ad esempio, il modulo semantico del modello Senso-Testo collega la rappresentazione semantica (RSem) di un enunciato a tutte le possibili rappresentazioni sintattiche profonde (RSintP) di quell’enunciato, il modulo sintattico profondo ha come input una rappresentazione sintattica profonda (RSintP) e “produce” come output tutte le possibili rappresentazioni sintattiche di superficie (RSintS), e così via (Milićević, 2006). Ciascun modulo opera dunque secondo il principio “un ingresso  $\leftrightarrow$  molte uscite” (Gatti, 1992:38). Ogni modulo prende il nome dal livello di rappresentazione che costituisce il suo input.

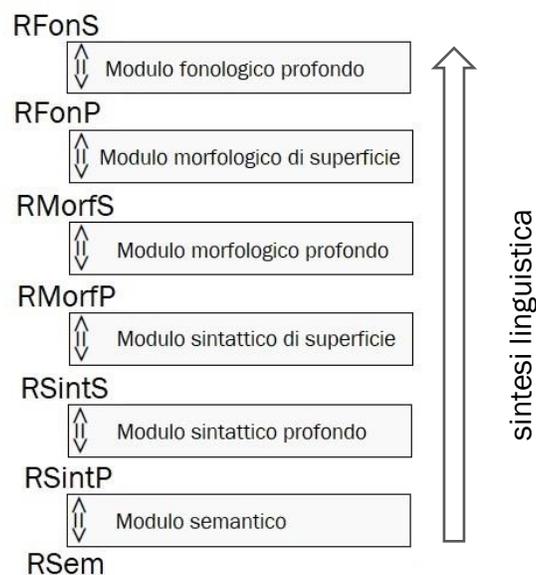


Figura 1: struttura del modello Senso-Testo (ripresa e adattata da Milićević, 2006).

Per riassumere il funzionamento globale del modello Senso-Testo possiamo utilizzare una metafora utilizzata dallo stesso Mel'čuk (2012), secondo cui il modello Senso-Testo è una catena di montaggio formata da diverse macchine, in cui a partire da materiale grezzo (il

senso, o meglio, la rappresentazione semantica) si arriva all'oggetto finito (il testo, o meglio la rappresentazione fonologica di superficie): ogni macchina ( $\approx$  un modulo) si occupa di uno stadio di lavorazione di un oggetto ( $\approx$  una rappresentazione) e lo passa poi al modulo successivo, dove l'oggetto subirà un'ulteriore elaborazione.

### 1.5.2 I livelli di rappresentazione linguistica del modello Senso-Testo

In questa sezione definiremo cosa si intende per *rappresentazione linguistica* <sup>$\alpha$</sup>  nell'approccio Senso-Testo e descriveremo i diversi livelli di rappresentazione linguistica, soffermandoci in particolare sui primi due: la rappresentazione semantica e la rappresentazione sintattica profonda.

In generale una rappresentazione di un enunciato a un determinato livello di analisi linguistica è definita come un insieme di oggetti formali chiamati *strutture* (Milićević, 2006). Tra queste è possibile distinguere una struttura centrale, che riflette le caratteristiche salienti dell'enunciato per il livello considerato, e una serie di strutture periferiche, portatrici di altre informazioni. Ad esempio, la rappresentazione semantica è costituita da una *struttura semantica* (che costituisce la struttura centrale) più una serie di altre strutture periferiche: la *struttura comunicativa*, la *struttura retorica*, e la *struttura referenziale* (cfr. sez. 1.5.2.1).

A ciascun livello di rappresentazione linguistica corrisponde un formalismo specifico di rappresentazione, un tipo di grafo diverso: la rappresentazione semantica (RSem) si avvale di *reti*, la rappresentazione sintattica profonda (RSintP) e quella di superficie (RSintS) si avvalgono di *alberi*, mentre le rappresentazioni di livello morfologico (RMorfP, RMorfS) e fonologico (RFonP, RFonS) utilizzano una forma di grafo più semplice, la *catena*. Come emerge dalla Figura 2, man mano che ci si avvicina alla superficie (ossia al testo, alla rappresentazione fonologica di superficie) la complessità dei grafi si riduce, e i grafi presentano un sempre maggior grado di linearizzazione.

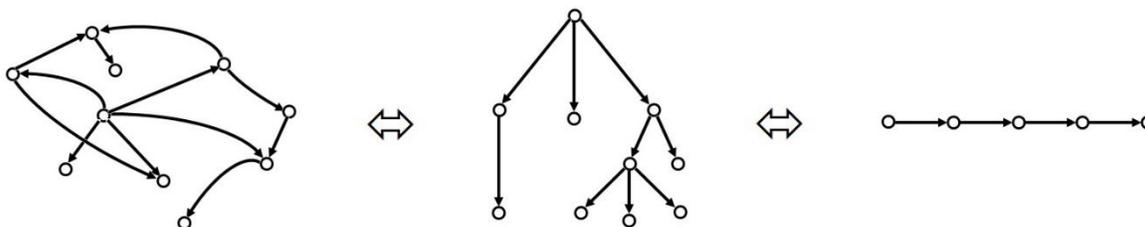


Figura 2: grafi del modello Senso-Testo: una rete, un albero e una catena (Gatti, 1992).

### 1.5.2.1 La rappresentazione semantica

La rappresentazione semantica (RSem) di un enunciato è un oggetto formale che rappresenta il **senso**<sup>α</sup> di tale enunciato, ossia il senso che il parlante vuole esprimere. Nella teoria Senso-Testo il senso è inteso unicamente come senso linguistico, escludendo aspetti pragmatici, culturali, ecc. Mel'čuk (2013) definisce il senso linguistico di un enunciato come la proprietà comune a tutte le **parafraresi**<sup>α</sup> di tale enunciato: la teoria Senso-Testo non offre dunque una definizione di senso, ma si serve della nozione primitiva di *equivalenza di senso*, poiché

il senso, in quanto oggetto dell'intuizione del parlante, non è accessibile direttamente all'osservazione né può essere manipolato da alcun modello linguistico. Si rende necessaria una sua trascrizione che dia luogo a rappresentazioni semantiche valide per tutte le lingue" (Rigotti, 1979:302).

Come anticipato nella sezione 1.5.2, la rappresentazione semantica è costituita da quattro strutture:

1. una *struttura semantica* (SSem, che costituisce la struttura centrale)
2. una *struttura semantico-comunicativa* (SSem-com, relativa alla distinzione tema/rema)
3. una struttura retorica (SRet, che fornisce informazioni sul registro e lo stile comunicativo adottato dal mittente: formale, colloquiale, ironico, poetico, ecc.).
4. una struttura referenziale (SRef, che specifica la corrispondenza tra gli elementi della struttura semantica e le entità del mondo reale, stabilendo una corrispondenza tra sensi e referenti).

Ai fini del presente lavoro ci concentreremo sulla prima struttura, e citeremo solo alcune caratteristiche della struttura comunicativa. Per approfondimenti si rimanda a Mel'čuk (2012: 183-413).

#### 1.5.2.1.1 La struttura semantica (SSem)

La struttura semantica consiste in una rete, che rappresenta il senso linguistico "puro" dell'enunciato.

I nodi del grafo corrispondono a unità di senso dette *semantemi*. Un **semantema**<sup>α</sup> è l'unità di base della struttura semantica e corrisponde al senso di un'unità lessicale (in termini generici,

al senso di una “parola”).<sup>12</sup> È possibile distinguere i semantemi in tre categorie: i predicati semantici, i nomi semantici e i quasi-predicati:

- un **predicato semantico**<sup>α</sup> è un semantema che denota una situazione che implica partecipanti (detti **attanti semantici**<sup>α</sup>), cioè una situazione in cui più elementi sono posti in relazione tra loro. Tipicamente, i predicati denotano eventi, azioni, processi, stati, percezioni, relazioni, proprietà, quantità, eccetera. I predicati semantici non si limitano ai verbi: anche molti sostantivi, a livello semantico, costituiscono dei predicati. Ad esempio il sostantivo *omicidio* ha due attanti: (1) l’omicida e (2) la vittima; il sostantivo *diritto* (cfr. Appendice 3: Scheda del lessema DIRITTO) ha ben quattro attanti: (1) il titolare del diritto, (2) l’oggetto del diritto, (3) l’autorità che conferisce il diritto e (4) i soggetti “altri” di fronte ai quali questo diritto viene fatto valere, e che sono tenuti a rispettarlo.
- un **nome semantico**<sup>α</sup> è un semantema che non implica nessun attante, e che in generale denota un’entità, come ad esempio nomi di individui, luoghi, date, sostanze.
- un **quasi-predicato**<sup>α</sup> è un semantema che, pur denotando un’entità, ha degli attanti. I quasi-predicati denotano entità che per natura (o per funzione) sono legate ad altre entità, poiché il loro stesso senso non può essere inteso se non in relazione ad altre entità. Sono esempi tipici di quasi-predicati: i nomi di professioni, come *professore* (persona x che insegna la materia y a z); i sostantivi che denotano relazioni personali o familiari come *sorella*, *padre*, *fidanzato* (perché implicano una relazione con qualcuno: una sorella è sorella *di qualcuno*, essere sorella non è una qualità intrinseca di una persona); sostantivi che denotano artefatti, cioè oggetti creati dall’uomo e spesso dotati di una funzione specifica, come *coltello* (oggetto che serve affinché una persona x tagli y);

---

<sup>12</sup> Oltre ai semantemi esistono altri tipi di unità di senso che possono essere incluse in una struttura semantica: i *grammemi* e i *quark semantici*. Tali unità si differenziano dai semantemi poiché non corrispondono al senso di unità lessicali, ma sono comunque portatrici di senso. Al fine di non complicare la presente esposizione i grammemi e i quark semantici sono stati omessi (anche dagli esempi): per approfondimenti si rimanda a Mel’čuk (2012: 56, 205).

I semantemi sono convenzionalmente identificati mediante una parola inclusa tra apici (ad esempio ‘portare’). Nella letteratura Senso-Testo (e nel Dizionario esplicativo-combinatorio) questa è la notazione utilizzata anche per indicare il senso di una parola o di un’espressione.

Gli archi del grafo corrispondono a due possibili tipi di relazioni: la *relazione predicato-attante* oppure la *relazione modificatore-semantema*.

Le relazioni tra un predicato semantico e i suoi attanti (ossia i suoi argomenti), sono rappresentate mediante archi orientati dal predicato verso l’attante e numerati (con numeri arabi) in base al grado dell’attante. Ad esempio, la struttura semantica (semplificata) dell’enunciato “Paolo regala un libro a Marina” sarebbe la seguente:

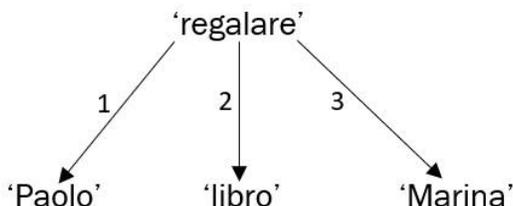


Figura 3: esempio di struttura semantica.

‘regalare’ è un predicato semantico, il suo senso denota una situazione che presuppone tre elementi essenziali (i suoi attanti semantici): 1) chi compie l’azione di regalare; 2) l’oggetto al centro dell’azione; 3) il destinatario dell’azione.<sup>13</sup> È importante sottolineare che nella struttura semantica gli attanti sono sempre logicamente ordinati, ossia l’attante 1 corrisponde sempre al primo argomento (l’agente), l’attante 2 corrisponde al secondo argomento, e così via. Ciascun arco che unisce il predicato ‘regalare’ a un suo attante è dunque etichettato con il numero dell’attante in questione.

L’altro tipo di relazione presente in una struttura semantica è la relazione tra un semantema e un altro semantema che funge da modificatore; un modificatore non è un attante (Kahane, 2003), bensì un semantema che contribuisce col proprio senso al senso del semantema a cui si riferisce. La relazione semantema-modificatore è rappresentata mediante un arco che va dal modificatore al semantema a cui esso si riferisce. Ad esempio, nella struttura

---

<sup>13</sup> Il numero e le caratteristiche degli attanti di un determinato lessema sono informazioni esplicitate nella voce del Dizionario esplicativo-combinatorio dedicata al lessema in questione (cfr. sez. 1.6 del presente capitolo).

semantica che rappresenta l'enunciato "Paolo regala un libro *antico* a Marina", la relazione tra il semantema 'libro' e il suo modificatore 'antico' sarebbe rappresentata con un arco che va da 'antico' a 'libro'.

Come già affermato, la struttura semantica rappresenta unicamente il senso dell'enunciato, ciò significa che, come è visibile in Figura 3, in una struttura semantica sono assenti tutti gli elementi non riguardanti il livello del senso: si fa totalmente astrazione della forma dell'enunciato omettendo preposizioni, connettori sintattici, indicatori della categoria grammaticale, forme flesse, verbi di supporto, ecc.

È inoltre opportuno specificare che due elementi che appartengono a due categorie grammaticali diverse ma hanno lo stesso senso (ad esempio il verbo *regalare* e il sostantivo *regalo*) nella struttura semantica corrispondono allo stesso semantema (in questo caso 'regalare').

Dato che il senso dell'enunciato è condiviso da tutte le parafrasi di tale enunciato, e che la struttura semantica rappresenta il puro senso di un enunciato, ne consegue che tutte le parafrasi di un enunciato hanno la medesima struttura semantica. In altre parole, riferendoci al processo di sintesi linguistica, un parlante che intende comunicare un determinato senso può scegliere tra vari modi in cui esprimerlo concretamente.

Riprendendo l'esempio già citato, tutti i seguenti enunciati condividono la struttura semantica riportata in Figura 3:

1. Paolo regala un libro a Marina
2. Il libro è un regalo di Paolo a Marina
3. Il regalo di Paolo a Marina è un libro
4. È a Marina che Paolo regala un libro

Le frasi hanno lo stesso senso, perché si riferiscono alla stessa situazione, in cui gli elementi coinvolti sono invariati. Ciò che differenzia queste frasi non è il senso, ma il modo di presentare l'informazione, ossia il risalto dato ad alcuni elementi. La struttura semantica non permette di rendere conto di queste differenze, esse sono rappresentate mediante un'altra struttura: la struttura semantico-comunicativa (SSem-com).

### 1.5.2.1.2 La struttura semantico-comunicativa (SSem-com)

La struttura semantico-comunicativa rappresenta le intenzioni comunicative del parlante, ossia il modo in cui egli intende “confezionare” il messaggio dando risalto ad alcuni elementi (Milićević, 2006). La SSem-com specifica dunque una serie di informazioni tra cui: qual è il *tema* (ciò di cui si parla) e qual è il *rema* (che cosa viene detto riguardo al tema); quali informazioni sono presentate come nuove e quali sono considerate note; quali elementi sono focalizzati (messi cioè in primo piano) eccetera. Tali informazioni sono dette *parametri semantico-comunicativi* (Milićević, 2006) e corrispondono alle *opposizioni semantico-comunicative* individuate da Mel'čuk (2012:323): *tematicità, novità, focalizzazione, prospettiva, enfasi, presupposizione, unitarietà, locuzionalità*.<sup>14</sup> Nella presente sezione per ragioni di spazio ci occuperemo solo della *tematicità* (opposizione tema/rema).

La SSem-com è una struttura che si sovrappone alla SSem e che dal punto di vista formale consiste non in un grafo, bensì in una divisione della SSem in diverse aree, ciascuna corrispondente a un valore di un parametro semantico-comunicativo: ad esempio la SSem-com specificherà quali aree della SSem corrispondono al tema (indicato come **T Sem**) e quali al rema (**R Sem**). In ciascuna delle aree così definite viene inoltre individuato un semantema principale, detto *nodo comunicativamente dominante*, che costituisce il fulcro dell'informazione di quell'area. Nella struttura Semantico-comunicativa i semantemi che costituiscono nodi comunicativamente dominanti sono segnalati con una sottolineatura (Figura 4 e Figura 5).

Dal momento in cui i parametri semantico-comunicativi sono definiti, il senso “puro” rappresentato dalla struttura semantica inizia ad acquisire una forma più precisa. Ciò significa che a parità di struttura semantica (cioè a parità di senso), due enunciati che dal punto di vista comunicativo sono organizzati diversamente avranno una struttura semantico-comunicativa diversa: si vedano gli esempi in Figura 4 e Figura 5.

---

<sup>14</sup> I termini italiani che designano le opposizioni semantico-comunicative sono stati da me conati a partire dalla loro versione francese in Milićević (2003) e nella tesi di Melissa Samson-Fauteux (2010), il cui relatore è lo stesso Mel'čuk.

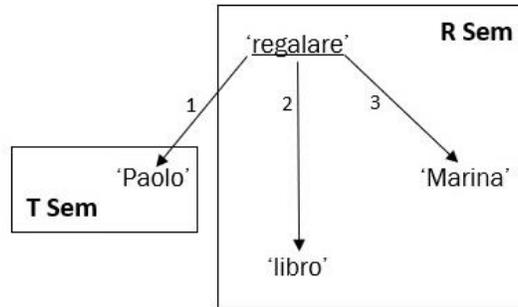


Figura 4: struttura semantico-comunicativa dell'enunciato "Paolo regala un libro a Marina".

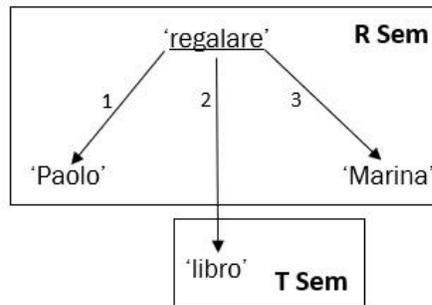


Figura 5: struttura semantico-comunicativa dell'enunciato "Il libro è un regalo di Paolo a Marina".

Nonostante la struttura semantico-comunicativa faccia parte delle cosiddette *strutture periferiche* della rappresentazione semantica, essa ha un ruolo tutt'altro che marginale: le informazioni di cui essa è portatrice vengono proiettate, attraverso il sistema di regole del modulo semantico (cfr. sezione 1.5.2.3), nella struttura sintattica profonda dell'enunciato. In pratica, la struttura semantico-comunicativa determina l'assetto sintattico dell'enunciato. La specificazione dei nodi comunicativamente dominanti nella SSem-com è cruciale per determinare la struttura sintattica profonda, perché (come vedremo nella sezione 1.5.3) il nodo comunicativamente dominante del rema all'interno della SSem-com costituisce il nodo-radice dell'albero sintattico profondo (Milićević, 2006). In termini generali, dal punto di vista del processo di sintesi linguistica ciò significa che una volta che il parlante decide a quali elementi di senso dare risalto, queste decisioni si riflettono sul modo in cui organizza l'enunciato.

### 1.5.2.2 La rappresentazione sintattica profonda

Passiamo ora alla descrizione del livello di rappresentazione successivo. Il passaggio dal livello di rappresentazione semantica al livello sintattico profondo avviene attraverso una componente del modello Senso-Testo detta *modulo semantico*, ossia un insieme di regole di corrispondenza che permettono di costruire la rappresentazione sintattica profonda a partire da una data rappresentazione semantica (cfr. sezione 1.5.1 e 1.5.2.3 del presente capitolo). Molte di queste regole si basano su regole a loro volta codificate nel lessico della lingua: il modello Senso-Testo infatti nel descrivere il funzionamento della lingua attribuisce un'importanza cruciale al lessico. Nel modello Senso-Testo è nel lessico che sono codificate informazioni sul contenuto semantico delle parole, ad esempio a quale situazione semantica una parola si riferisce, oppure la sua struttura di attanti semantici; è nel lessico che sono codificate anche alcune informazioni sul comportamento sintattico delle parole, ad esempio quanti argomenti sintattici ha un verbo. La trattazione del lessico nel modello Senso-Testo sarà ripresa e approfondita nella sezione 1.6 del presente capitolo. Per ora basti sapere che il lessico, per le informazioni di cui è portatore, ha un ruolo cruciale nel passaggio dal livello semantico al livello sintattico profondo, e che le unità di base della rappresentazione sintattica profonda sono, non a caso, unità *lessicali*.

Il formalismo di cui si avvale il modello Senso-Testo per il livello sintattico (sia il livello profondo che il livello di superficie) è la rappresentazione mediante alberi a dipendenze: la struttura dell'enunciato è rappresentata non in termini di costituenti (come in sintassi generativa, ad esempio), bensì in termini di relazioni binarie di dipendenza tra gli elementi che compongono l'enunciato (Milićević, 2006).

Anche la rappresentazione sintattica profonda è costituita da un insieme di quattro strutture: una struttura centrale, ossia la *struttura sintattica profonda*, e tre strutture periferiche: una *struttura comunicativa*, una *struttura prosodica* e una *struttura anaforica*. Per ragioni di spazio nel presente lavoro si tratterà solo della struttura sintattica profonda. Per approfondimenti si rimanda a Mel'čuk (2013).

La struttura sintattica profonda è un albero a dipendenze in cui i nodi corrispondono a *unità lessicali profonde* (convenzionalmente riportate in caratteri maiuscoli), una categoria strutturale che comprende tre diversi tipi di elementi:

1. unità lessicali piene, ulteriormente distinguibili in lessemi e locuzioni;
2. funzioni lessicali;
3. *lessemi "fittizi"*, ossia costruzioni sintattiche specifiche portatrici di senso, ma che non corrispondono a lessemi che compaiono in sintassi di superficie; per ragioni di spazio e per non complicare l'esposizione essi sono esclusi dalla presente trattazione, rimandando il lettore a Mel'čuk (2013).

L'*unità lessicale*<sup>α</sup> *piena* in linguistica Senso-Testo è una categoria lessicale che comprende i lessemi e le locuzioni idiomatiche, ossia le unità di base del lessico di una lingua naturale. Un *lessema*<sup>α</sup> è definibile in termini generici come la "forma base" di una parola (cioè la forma non flessa, quella che tipicamente costituisce l'entrata di una voce in un dizionario), dotata di un significato lessicale. Nella struttura sintattica profonda dunque le parole non sono riportate nella forma (flessa) in cui compaiono nell'enunciato, bensì nella loro forma base.<sup>15</sup> Sarà solo al livello successivo, il livello sintattico di superficie, che le parole saranno riportate nella loro forma flessa.

È necessario specificare che nella struttura sintattica profonda compaiono solo *lessemi semanticamente pieni* (ossia dotati di uno specifico senso).<sup>16</sup>

Ogni lessema è accompagnato da alcune informazioni grammaticali semanticamente rilevanti, dette *grammemi semanticamente pieni*, che corrispondono cioè a un significato che il mittente desidera *intenzionalmente* esprimere (distinzioni come definito/indefinito per le lingue in cui tale distinzione è rilevante, la distinzione singolare/plurale per i sostantivi, e l'indicazione su voce, modo e tempo per i verbi). I grammemi semanticamente pieni sono riportati in pedice in lettere minuscole. Viceversa, le informazioni grammaticali che non costituiscono *grammemi semanticamente pieni* bensì *grammemi sintattici* (Kahane, 2003:12), come il genere e il caso dei sostantivi, il genere, numero e caso degli aggettivi, o ancora la persona del verbo, non sono specificati a livello sintattico profondo, ma compaiono

---

<sup>15</sup> Si noti la differenza rispetto alla struttura semantica, nella quale non vengono riportate le parole, bensì solo il loro senso, dunque senza alcuna distinzione tra verbi e sostantivi aventi un medesimo senso. Nella struttura sintattica profonda gli elementi compaiono nella loro forma base, in cui la categoria grammaticale (verbo, sostantivo, aggettivo, avverbio) è definita.

<sup>16</sup> I lessemi *semanticamente vuoti*, ad esempio alcune preposizioni, congiunzioni, verbi di supporto e altre cosiddette parole grammaticali non fanno parte di questo livello di descrizione: essi compaiono nella struttura sintattica di superficie.

nel livello successivo, nella rappresentazione sintattica di superficie. Il motivo è che le informazioni di questo tipo “non fanno parte di ciò che il mittente *desidera* esprimere; esse rappresentano piuttosto *ciò che la lingua richiede al mittente di esprimere*” (Kahane, 2003:8, traduzione e corsivo miei).

Una **locuzione**<sup>α</sup> (detta anche locuzione idiomatica) è un’espressione fissa, formata da più parole, il cui significato non è costruito a partire dalla somma dei significati delle parole che la compongono – bensì è dotata di un significato proprio e specifico (Cfr. cap. 2 sez 2.3.2). Nella struttura sintattica profonda una locuzione idiomatica, che costituisce un unico “blocco” di senso, occupa un unico nodo, non è scomposta.

Un esempio di quanto esposto finora, riportato in Figura 6, è la struttura sintattica profonda dell’enunciato “I giornali italiani trovano un capro espiatorio”:

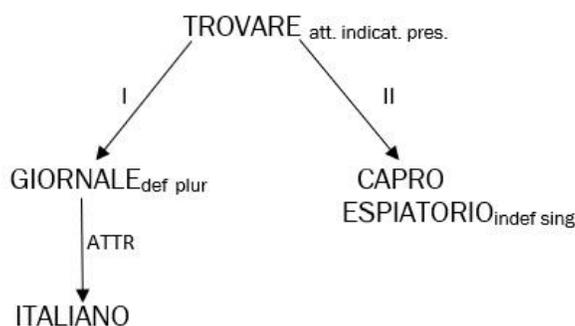


Figura 6: esempio di struttura sintattica profonda (SSintP).

“I giornali” viene riportato nella sua forma base, ossia il lessema GIORNALE, con il grammema che indica il plurale; l’aggettivo “italiani” è rappresentato dal lessema ITALIANO, senza indicazioni grammaticali (poiché genere e numero degli aggettivi non sono specificati nella struttura sintattica profonda); il verbo “trovano” è rappresentato dal lessema TROVARE, in cui sono specificati i grammemi di voce (attiva), modo (indicativo) e tempo (presente); CAPRO ESPIATORIO è una locuzione idiomatica e pertanto occupa un solo nodo, non è scisso nelle sue componenti (sostantivo + aggettivo).

Una **funzione lessicale**<sup>α</sup> è un tipo particolare di unità lessicale profonda: nella struttura sintattica profonda le funzioni lessicali sono utilizzate per rappresentare le collocazioni, ossia espressioni formate da più parole in cui uno degli elementi (detto base della collocazione) è scelto liberamente dal locutore per esprimere un senso, mentre la scelta degli altri elementi

(i collocati della base) è condizionata, oltre che in funzione del senso da esprimere, da restrizioni lessicali particolari (cfr. cap. 2, sez. 2.3.2).

In particolare, nella struttura sintattica profonda la base della collocazione è rappresentata dal lessema corrispondente, e occupa un nodo dell'albero, mentre il collocato della base è rappresentato non da un lessema, bensì dal nome della funzione lessicale a cui esso corrisponde, in maiuscoletto, e occupa un nodo distinto dell'albero.

Per fare un esempio presentiamo la struttura sintattica profonda dell'enunciato "I giornali muovono dure critiche al governo" (Figura 7); per una trattazione più approfondita delle funzioni lessicali si rimanda al cap. 2 sez. 2.4 e sez. 2.5. "Dure critiche" è una collocazione la cui base è "critiche", mentre "dure" è un collocato che può essere rappresentato mediante la funzione lessicale MAGN (una funzione lessicale il cui senso, come vedremo nel capitolo 2, è "intensità"). Allo stesso modo "muovere critiche" è una collocazione in cui la base è "critiche", mentre "muovere" è un collocato che sintatticamente svolge la funzione di verbo di supporto: "muovere" è dunque la radice dell'albero sintattico (in quanto verbo), ed è rappresentato mediante la funzione lessicale OPER<sub>1</sub>.

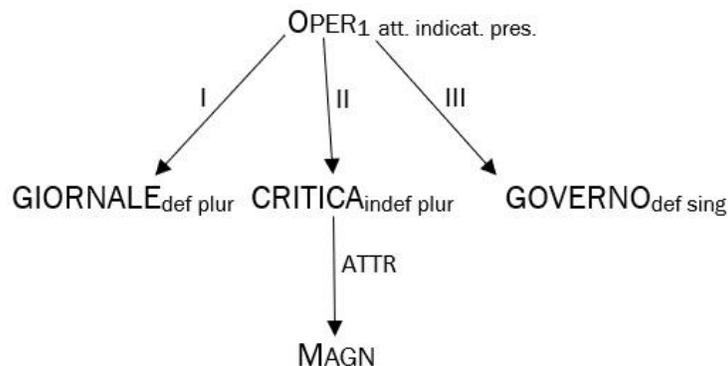


Figura 7: struttura sintattica profonda dell'enunciato "I giornali muovono dure critiche al governo"

Nella struttura sintattica profonda le funzioni lessicali occupano il nodo che dovrebbe essere occupato dal collocato, dunque, per così dire, lo "sostituiscono". Il collocato comparirà solo nel passaggio al livello successivo, ossia nella struttura semantica di superficie.

Dopo aver descritto gli elementi che costituiscono i nodi della struttura sintattica profonda presenteremo brevemente gli archi del grafo. Gli archi della struttura sintattica profonda rappresentano relazioni di dipendenza tra unità lessicali profonde. Esistono 4 macro-tipi di

relazioni sintattiche profonde, ma in questa sede ci concentreremo in particolare sui due tipi principali:

- la *relazione attanziale* è la relazione tra un elemento e un suo attante semantico profondo. Gli **attanti sintattici profondi**<sup>α</sup> corrispondono tipicamente agli argomenti sintattici di un verbo o di un'altra unità lessicale. La relazione tra un'unità lessicale profonda e un suo attante è rappresentata mediante un arco orientato dal predicato verso l'attante (poiché l'attante dipende dal predicato) e numerato (in numeri romani) in base al grado dell'attante. L'esempio in Figura 7 presenta la costruzione verbale "X muove una critica a Y" in cui il verbo "muovere" ha tre attanti sintattici profondi: (I) chi "muove", (II) che cosa "è mosso", (III) chi è il destinatario. In genere nelle lingue europee il primo attante sintattico corrisponde al soggetto, il secondo al complemento oggetto dei verbi transitivi alla forma attiva, ecc.
- la *relazione attributiva* è la relazione tra un elemento e un suo modificatore (in genere un attributo). All'interno della struttura sintattica profonda la relazione attributiva è rappresentata mediante un arco orientato da un elemento al suo modificatore e etichettato mediante la notazione ATTR in maiuscolo. Nell'esempio in Figura 6 la relazione attributiva è quella che lega il lessema GIORNALE al lessema ITALIANO; nell'esempio in Figura 7 è presente una relazione attributiva che lega il lessema CRITICA alla funzione lessicale MAGN, la quale rappresenta un attributo portatore del senso "intensità".

Gli altri due tipi di relazioni sintattiche sono la *relazione appenditiva* e la *relazione coordinativa*, che qui accenniamo solo brevemente. La relazione appenditiva lega il verbo principale a elementi "extra strutturali" della frase come ad esempio interiezioni, allocutivi, ecc. (Milićević, 2006); essa è rappresentata da un arco recante l'etichetta APPEND. La *relazione coordinativa* esprime invece la coordinazione sintattica tra elementi del grafo ad albero (ad esempio la coordinazione tra i due soggetti della frase "Marco e Luca frequentano il liceo scientifico"); essa è rappresentata da un arco recante l'etichetta COORD.

Le relazioni sintattiche profonde sono state ideate per essere universali, cioè per essere in grado di rappresentare le relazioni sintattiche di qualunque lingua naturale.

### 1.5.2.3 Dalla RSem alla RSintP: il modulo semantico

Come anticipato all'inizio della sezione 1.5.2.2, il passaggio da rappresentazione semantica a rappresentazione sintattica profonda avviene attraverso un insieme di regole, parte del cosiddetto modulo semantico, che a partire dalla RSem permette di costruire la RSintP. Di seguito presentiamo sinteticamente le regole del modulo semantico, suddivise in regole di corrispondenza e regole di equivalenza. Alcune regole saranno solo citate senza essere spiegate. Per approfondimenti si rimanda a Mel'čuk (2013).

Le *regole di corrispondenza* sono a loro volta divise in vari tipi (di cui approfondiamo solo i primi tre):

- 1) *regole di lessicalizzazione*: stabiliscono le corrispondenze tra semantemi (i nodi del grafo semantico) e unità lessicali profonde (i nodi dell'albero sintattico profondo). In altre parole le regole di lessicalizzazione "costruiscono" i nodi dell'albero sintattico a partire dai nodi del grafo semantico: permettono di arrivare dai semantemi alle unità lessicali profonde, stabilendo ad esempio se il lessema 'regalare' verrà reso mediante il lessema REGALARE oppure mediante il lessema REGALO. Le regole di lessicalizzazione sono basate su informazioni codificate nel lessico della lingua (cfr. sez. 1.6);
- 2) *regole di morfologizzazione*: stabiliscono le corrispondenze tra i grammemi del grafo semantico e i valori grammaticali associati alle unità lessicali dell'albero sintattico; in pratica le regole di morfologizzazione controllano la resa delle informazioni grammaticali nell'albero sintattico;
- 3) *regole di arborizzazione*: si tratta delle regole che stabiliscono la struttura dell'albero sintattico profondo, determinando le dipendenze sintattiche profonde a partire dalle corrispondenti relazioni semantiche e semantico-comunicative. In altre parole le regole di arborizzazione permettono di trasformare le relazioni della struttura semantica nelle corrispondenti relazioni sintattiche profonde, trasformando così la rete semantica in albero sintattico. Un esempio è la regola che trasforma una relazione modificatore-semantema presente nella struttura semantica (un arco orientato dal modificatore al semantema a cui esso si riferisce) in una relazione sintattica *attributiva* (un arco orientato nel senso opposto, che va dall'unità lessicale profonda al suo

modificatore. Cfr. esempi in sez. 1.5.3). Anche le regole di arborizzazione dipendono da informazioni codificate nel lessico della lingua.

Oltre ai tre tipi di regole elencati esistono anche le *regole comunicative* (che costruiscono la struttura comunicativa a livello sintattico profondo) e le *regole prosodiche* (che danno origine alla struttura prosodica della rappresentazione sintattica profonda), per la cui trattazione si rimanda a Milićević (2006) e Kahane (2003).

Le *regole di equivalenza* del modulo semantico, dette anche *regole parafrastiche*, sono invece divise in due tipi:

- 1) *regole di equivalenza semantica*, che stabiliscono l'equivalenza tra due rappresentazioni semantiche (o porzioni di esse);
- 2) *regole di equivalenza lessicale e sintattica*, che stabiliscono l'equivalenza tra due rappresentazioni sintattiche profonde e permettono di costruire rappresentazioni sintattiche equivalenti tra loro (ossia che esprimono lo stesso senso) ma organizzate diversamente. Un esempio classico è l'equivalenza tra il lessema CRITICARE e la funzione lessicale  $OPER_1(CRITICA)=muovere\ una\ critica$ . Le regole di equivalenza lessicale e sintattica si distinguono dalle regole di lessicalizzazione poiché queste ultime servono a “costruire” i nodi dell'albero sintattico profondo a partire dai nodi della rappresentazione semantica, mentre le regole di equivalenza lessicale e sintattica controllano i meccanismi di parafrasi che operano a livello sintattico profondo.

### **1.5.3 Esempio di sintesi linguistica: dal livello semantico al livello sintattico profondo**

Chiudiamo la sezione sul modello Senso-Testo presentando un esempio concreto che chiarisca e aiuti a fissare i concetti esposti finora: useremo il modello Senso-Testo per rappresentare le prime due tappe della sintesi linguistica di due enunciati. Mostreremo come, a partire da un senso che il locutore desidera esprimere, si possa arrivare a due possibili rappresentazioni semantiche e poi a due possibili rappresentazioni sintattiche profonde, a partire dalle quali sarebbe poi possibile continuare la sintesi linguistica attraverso il livello sintattico di superficie, il livello morfologico e quello fonologico, per arrivare agli enunciati nella loro forma testuale. In ogni caso nel presente lavoro ci fermeremo al livello sintattico profondo). L'esempio è tratto da Milićević (2006), adattato all'italiano e semplificato.

Immaginiamo la seguente situazione di senso: il governo ha deciso di aumentare le tasse, e i giornali lo criticano per questa decisione. A livello semantico, questa situazione sarebbe rappresentata mediante la seguente struttura semplificata:<sup>17</sup>

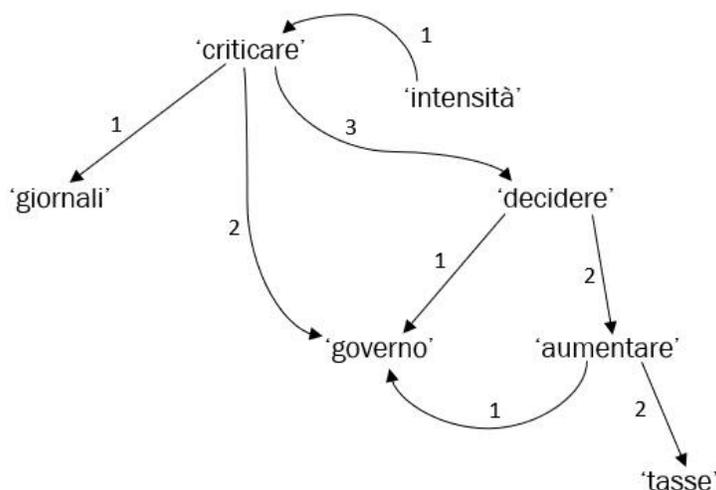


Figura 8: esempio di struttura semantica (SSem). Adattata da Milićević (2006).

Il semantema 'criticare' ha 3 attanti semantici: chi critica (1), chi è criticato (2), e il motivo (3) ("X critica Y per Z"). Il semantema 'intensità' è un modificatore di 'criticare' e non un suo attante, pertanto la relazione che li unisce è un arco che va da 'intensità' a 'criticare', e non viceversa. Il semantema 'decidere' è il terzo attante di 'criticare', e a sua volta ha due attanti poiché esprime una situazione di senso in cui qualcuno (1) decide qualcosa (2). 'governo' è il primo attante di 'decidere' e anche il primo attante di 'aumentare', semantema il cui secondo attante (la cosa che viene aumentata) è 'tasse'.

Ricordiamo che i semantemi sono unità di senso, non parole, dunque il fatto che alcuni di essi siano rappresentati mediante verbi è solo una convenzione. È a livello sintattico profondo che si definisce se tale senso sarà espresso con un verbo (il lessema CRITICARE) o con un sostantivo (il lessema CRITICA).

<sup>17</sup> In questa struttura semantica (e nelle strutture semantiche successive) sono stati omessi i grammemi, e alcuni semantemi sono al plurale appunto per evitare di dover riportare i grammemi 'singolare' e 'plurale', come dovrebbe avvenire in teoria. Inoltre i semantemi che esprimono un senso complesso come 'decidere' non sono stati ulteriormente scomposti in semantemi più semplici, come normalmente si dovrebbe fare.

Come illustrato nella sezione 1.5.2.1.2, un locutore può decidere di dare risalto ad alcuni elementi di senso: questa preminenza è marcata mediante la struttura comunicativa, la struttura che si sovrappone alla struttura semantica dividendola in due aree, corrispondenti rispettivamente al tema (T Sem) e al rema (R Sem). Di seguito presentiamo due possibili configurazioni tema-rema della struttura semantico-comunicativa.

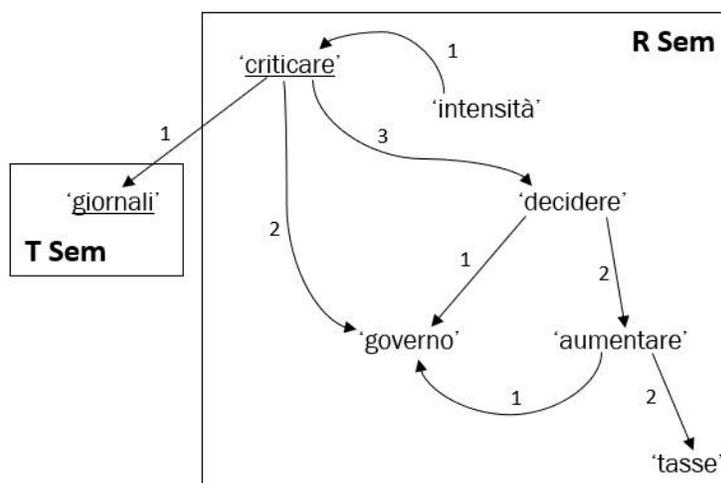


Figura 9: esempio di struttura semantico-comunicativa (Ssem-com).

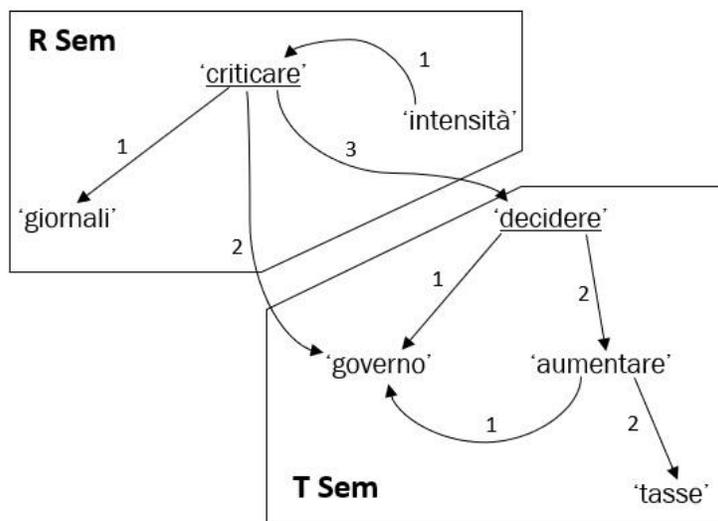


Figura 10: esempio di struttura semantico-comunicativa (Ssem-com).

Nell'esempio in Figura 9 il tema (ciò di cui si parla) è 'giornali', mentre il rema (ciò che viene affermato a proposito del tema) è tutto il resto, ossia il fatto che essi criticano il governo per la decisione di aumentare le tasse. I nodi dominanti del tema e del rema sono sottolineati.

Guardando questa struttura comunicativa possiamo intuire quali frasi esprimono questa divisione tema/rema. Eccone una:

- (1) I giornali hanno duramente criticato il governo per la sua decisione di aumentare le tasse.

Nell'esempio in Figura 10 il tema è invece la decisione del governo di aumentare le tasse, mentre il rema è tutto il resto, ossia il fatto che questa decisione sia criticata dai giornali. Il nodo dominante del tema (sottolineato) in questo caso è 'decidere', dunque un esempio di frase che può esprimere questo assetto è:

- (2) La decisione del governo di aumentare le tasse ha ricevuto dure critiche da parte dei giornali.

Sappiamo poi che la struttura semantico-comunicativa influisce sull'assetto sintattico profondo dell'enunciato in formazione: il nodo dominante del rema nella struttura semantico-comunicativa diventa il nodo-radice nella struttura sintattica profonda (cioè il verbo principale), mentre il nodo dominante del tema in genere diventa il primo attante sintattico del verbo principale (cioè il soggetto sintattico del periodo). Il modo in cui viene a definirsi il resto della struttura sintattica (l'assetto di nodi e relazioni sintattiche) è controllato dalle regole del modulo semantico esposte nella sezione precedente (che qui, per ragioni di spazio, non saranno riprese).

Da due rappresentazioni semantiche che si distinguono a livello semantico-comunicativo si possono dunque ottenere rappresentazioni sintattiche diverse, che proseguendo nel processo di sintesi linguistica daranno origine a enunciati diversi. La Figura 11 e la Figura 12 ne sono un esempio.

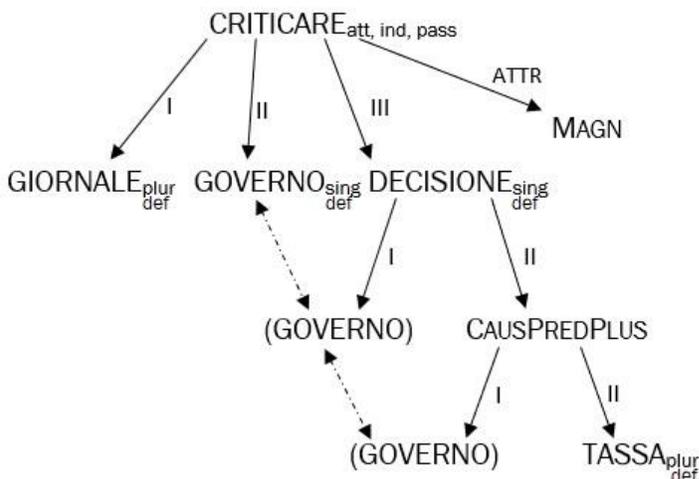
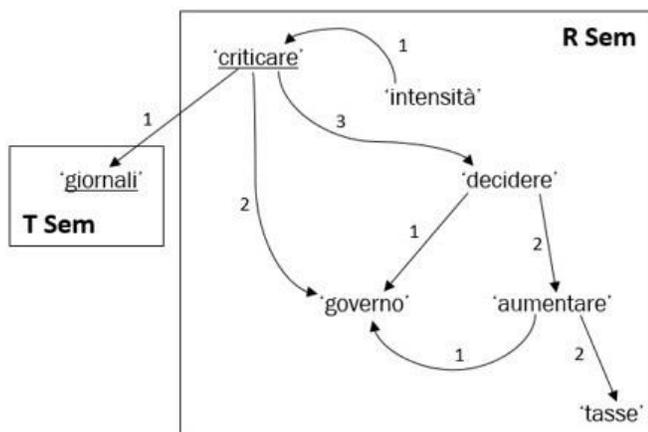


Figura 11: SSem + SSem-com (a sinistra) e SSintP (a destra) dell'enunciato (1) "I giornali hanno duramente criticato il governo per la sua decisione di aumentare le tasse".

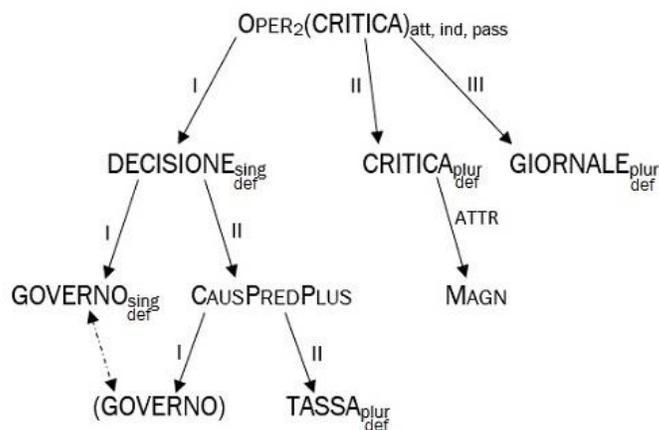
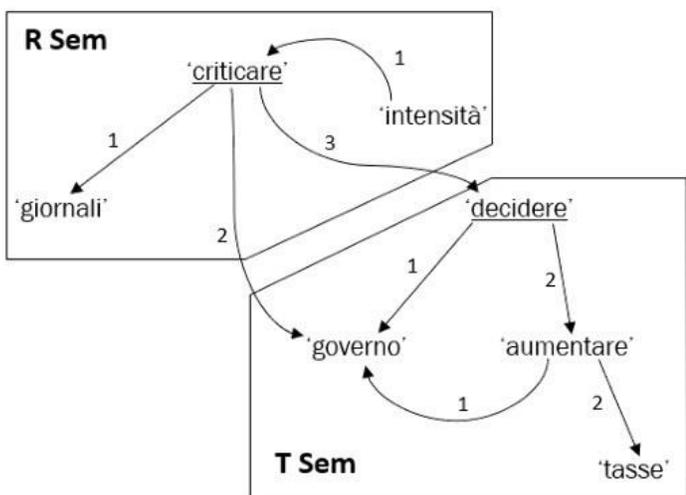


Figura 12: SSem + SSem-com (a sinistra) e SSintP (a destra) dell'enunciato (2) "La decisione del governo di aumentare le tasse ha ricevuto dure critiche da parte dei giornali".

L'esempio in Figura 11 mostra una struttura semantico-comunicativa in cui il nodo dominante del tema è 'giornali', mentre il nodo dominante del rema è 'criticare'. Dunque 'criticare' nella struttura sintattica profonda diventerà il verbo principale di tutto il periodo, il nodo-radice dell'albero, ossia il lessema CRITICARE. Il semantema 'giornali' diventa invece il primo attante sintattico profondo del verbo principale, ossia il suo soggetto sintattico. Gli altri attanti sintattici profondi di CRITICARE sono GOVERNO (un lessema derivato dal semantema 'governo' mediante una regola di lessicalizzazione) e DECISIONE (anch'esso prodotto di una

regola di lessicalizzazione che determina la sua realizzazione come sostantivo). Un altro elemento che dipende da CRITICARE, ma che non è un attante, è la funzione lessicale MAGN(CRITICARE), legata a CRITICARE da una relazione attributiva (ATTR). Questo nodo dell'albero è il prodotto di una regola di lessicalizzazione particolare, che trasforma il senso 'intensità' della struttura semantica non in un lessema, bensì in un altro tipo di unità lessicale profonda: una funzione lessicale. Una regola di lessicalizzazione analoga trasforma 'aumentare' nella funzione lessicale CAUSPREDPLUS, funzione lessicale verbale complessa (cfr. sez. 2.4.2.3) il cui senso è 'causare l'incremento di qualcosa'.

DECISIONE è un lessema che possiede due attanti sintattici, che corrispondono agli attanti semantici di 'decidere'. Si noti che il primo attante è sempre GOVERNO, che in questo caso viene ripetuto (perché non è possibile tracciare archi ascendenti in un grafo ad albero, cosa che invece era possibile fare nella rete semantica). I nodi del grafo occupati da GOVERNO sono collegati tra loro da archi bidirezionali a linea tratteggiata: tali archi non fanno parte della struttura sintattica, bensì costituiscono la *struttura anaforica* della rappresentazione sintattica profonda (cfr. sez. 1.5.2.2). La struttura anaforica rappresenta i legami di co-referenzialità tra nodi della struttura sintattica profonda e fornisce le informazioni necessarie per comprendere le operazioni di pronominalizzazione ed ellissi, che di norma vengono realizzate in una fase successive della sintesi linguistica (Milićević, 2006).

Nell'esempio in Figura 11 il lessema CRITICARE è parte di una costruzione sintattica a tre attanti (X critica Y per Z), ma è possibile trovare CRITICARE in costruzioni a due attanti (X critica Z) (Cfr. sez. 1.6.1.2).

L'esempio in Figura 12 mostra invece una struttura semantico-comunicativa in cui il tema non è un nodo singolo, ma un'intera porzione di grafo, in cui il nodo dominante è 'decidere'. Il nodo dominante del rema è 'criticare'. A causa di questa diversa configurazione, nel passaggio dal livello semantico al livello sintattico profondo avviene qualcosa di molto diverso dall'esempio in Figura 11: questa volta il primo attante sintattico del verbo principale (ossia il suo soggetto sintattico) è un'unità lessicale che corrisponde al senso 'decidere', ossia il sostantivo DECISIONE. Il verbo principale dev'essere un'unità lessicale profonda che corrisponda al senso di 'criticare' (poiché 'criticare' è il nodo dominante del rema), ma c'è un problema: il verbo CRITICARE non ammette DECISIONE come primo attante sintattico, perché la decisione è la cosa che viene criticata. Come trovare un verbo adeguato? In questo caso interviene una

regola di lessicalizzazione che determina la realizzazione di ‘criticare’ sotto forma di funzione lessicale, in modo che nella fase successiva della sintesi linguistica sia possibile selezionare un verbo adeguato al contesto lessicale e sintattico. La funzione lessicale, in questo caso, è OPER<sub>2</sub>(CRITICA), che corrisponde al verbo “ricevere (critiche)”. OPER<sub>2</sub>(CRITICA) è dunque il nodo radice dell’albero sintattico profondo, che ha come primo attante DECISIONE, come secondo attante CRITICA e come terzo attante GIORNALE (= “la decisione ha ricevuto critiche da parte dei giornali”).

Abbiamo menzionato il fatto che il verbo corrispondente alla funzione lessicale OPER<sub>2</sub> viene selezionato solo nel passaggio al livello successivo: questo perché nella struttura sintattica profonda una funzione lessicale occupa il nodo che sarebbe riservato a una parola, e questa parola compare solo nella struttura di livello successivo, ossia la struttura sintattica di superficie, grazie a una regola di corrispondenza del modulo successivo.

Concludiamo qui la spiegazione delle prime due fasi della sintesi linguistica. Lo scopo di questa sezione era fare un esempio concreto che aiutasse a capire i primi due livelli di rappresentazione (semantica e sintattica profonda), e iniziare a introdurre il ruolo delle funzioni lessicali nel modello Senso-Testo. Per approfondimenti si rimanda a Polguère (1998), Kahane (2003), Milićević (2006) e, ovviamente, a Mel’čuk (2012, 2013).

## **1.6 Il ruolo del lessico nell’approccio Senso-Testo: il Dizionario esplicativo-combinatorio (DEC)**

In questa sezione presenteremo sinteticamente l’approccio Senso-Testo al lessico e il modo in cui il lessico è descritto e formalizzato.

Nell’approccio Senso-Testo la descrizione di una lingua naturale si basa su due componenti: la grammatica e il lessico.

- La **grammatica** di una lingua è “l’insieme di tutti i segni grammaticali e di tutte le regole di tale lingua” (Mel’čuk 2013:260). In termini generici potremmo dire che le regole contenute nei diversi moduli del modello Senso-Testo corrispondono alla grammatica di una lingua.

- Il **lessico** è l'insieme di tutti i *segni lessicali* di una lingua, ossia delle unità lessicali (lessemi e locuzioni) e delle relative descrizioni (Mel'čuk 2013:260). In altre parole, nell'approccio Senso-Testo il lessico non è solo un insieme di unità lessicali ma contiene anche le descrizioni di tali unità lessicali, ossia informazioni sulle loro proprietà semantiche, sintattiche e combinatorie.

Nel modello Senso-Testo il lessico ha un ruolo cruciale nel funzionamento della lingua, in particolare svolge un ruolo di interfaccia tra il livello semantico e il livello sintattico profondo, poiché le informazioni in esso codificate si riflettono nelle regole del modulo semantico (in particolare nelle regole di arborizzazione e di lessicalizzazione, cfr. sez. 1.5.2.3).

Lo strumento formale attraverso cui, nel modello Senso-Testo, il lessico è descritto in maniera sistematica è il Dizionario esplicativo-combinatorio (d'ora in avanti abbreviato in DEC). Il DEC è uno strumento formale nel senso che è "parte integrante del modulo semantico del modello Senso-Testo" (Mel'čuk 2013:261), poiché esso contiene le informazioni su cui sono basate le regole del modulo semantico.

Come suggerito dal suo stesso nome, il Dizionario esplicativo-combinatorio è anche un vero e proprio *dizionario* monolingue contenente tutte le informazioni linguistiche relative a ogni unità lessicale (significati, struttura di attanti, uso in contesto, sinonimi, derivati, ecc.): da qui l'aggettivo *esplicativo*; infine, come indicato dall'aggettivo *combinatorio*, esso fornisce una descrizione esaustiva delle proprietà combinatorie delle unità lessicali, ossia le loro co-occorrenza con specifiche unità lessicali a cui esse sono legate da relazioni lessicali.

Non si tratta di un dizionario classico, inteso per la consultazione generale, bensì un catalogo di informazioni linguistiche creato per scopi primariamente teorici: avere una descrizione dettagliata e sistematica del lessico di una lingua secondo i principi della teoria Senso-Testo. Il DEC utilizza un metalinguaggio basato su concetti, terminologia e schemi propri dell'approccio Senso-Testo: anche per questo motivo si configura come un oggetto destinato allo studio e alla ricerca più che alla consultazione da parte di un pubblico non specializzato. Ad esempio, per la sua ricchezza di informazioni relative alle collocazioni linguistiche il DEC può rivelarsi uno strumento utile nella didattica delle lingue e nella formazione di traduttori e interpreti: da almeno un decennio gli strumenti lessicografici come il DEC sono oggetto di studio e di applicazioni in didattica presso l'Université de Montréal, grazie al lavoro

dell'Observatoire Linguistique Sens-Texte (Polguère, 2000, 2003; Anctil, 2005; Mel'čuk e Polguère, 2007; Tremblay, 2009; L'Homme, 2010, Frassi e Tremblay, 2016).

I DEC cartacei esistenti sono stati elaborati per la lingua russa (Mel'čuk e Žolkovskij, 1984) e per il francese (Mel'čuk e al., 1984, 1988, 1992, 1999; Mel'čuk e Polguère, 2007). Il DiCo, database lessicale della lingua francese ispirato al DEC (Cfr. sez. 1.3.2) è consultabile anche tramite l'interfaccia online DiCouèbe.<sup>18</sup> Esiste poi un interessante progetto lessicografico per la lingua francese portato avanti da Alain Polguère e altri ricercatori del laboratorio ATILF:<sup>19</sup> il Réseau Lexical du Français.<sup>20</sup> Esiste inoltre un dizionario delle collocazioni della lingua spagnola costruito secondo i criteri della lessicologia esplicativa-combinatoria: il DICE (*Diccionario de colocaciones del Español*)<sup>21</sup> elaborato da Margarita Alonso Ramos a partire dal 2003. Segnaliamo inoltre il lavoro di Marie-Claude L'Homme, che ha applicato i principi della lessicologia esplicativa-combinatoria allo studio della terminologia elaborando i due database terminologici DiCoEnviro (relativo al dominio dell'ambiente)<sup>22</sup> e DiCoInfo (relativo al dominio dell'informatica).<sup>23</sup>

Ad oggi non esistono opere lessicografiche analoghe per la lingua italiana. Gli esempi presentati di seguito sono pertanto ripresi dalla versione francese.

### 1.6.1 Struttura di una voce del DEC

Come già anticipato nella sezione 1.5.2.2 del presente capitolo, nella teoria Senso-Testo è nel lessico che sono codificate informazioni semantiche, sintattiche e relative alle proprietà combinatorie dei lessemi. Ecco perché ciascuna voce del DEC è articolata in tre parti principali (Milićević, 2006): la parte semantica, la parte sintattica<sup>24</sup> e la parte relativa alle relazioni lessicali, che presenteremo in estrema sintesi.

---

<sup>18</sup> <http://olst.ling.umontreal.ca/dicouebe/>

<sup>19</sup> <http://www.atilf.fr/>

<sup>20</sup> <http://www.atilf.fr/spip.php?rubrique604&idfirst=3833>

<sup>21</sup> <http://www.dicesp.com/paginas>

<sup>22</sup> [http://olst.ling.umontreal.ca/?page\\_id=214](http://olst.ling.umontreal.ca/?page_id=214)

<sup>23</sup> [http://olst.ling.umontreal.ca/?page\\_id=53](http://olst.ling.umontreal.ca/?page_id=53)

<sup>24</sup> Nel DEC francese questa parte è chiamata *régime*, ma per il termine italiano ci siamo basati sul termine inglese *syntactic zone*, attestato in Kahane (2003) e Milićević (2006).

### 1.6.1.1 Parte semantica

Questa parte contiene la definizione dell'unità lessicale. Tale definizione consiste in una scomposizione semantica del senso di un'unità lessicale in sensi più semplici, detti primitivi semantici, sul modello della definizione aristotelica per genere prossimo e differenza specifica (cfr. **definizione**<sup>α</sup>).

La definizione è in genere accompagnata da una *forma proposizionale*, ossia una formula che permette di esplicitare gli attanti che l'unità lessicale possiede (Jousse e Polguère, 2005). Un esempio è la prima tra le possibili definizioni del lessema CRITICARE (tratta da Mel'čuk e al., 1999:207, traduzione mia): "X esprime un'opinione sfavorevole su Y riguardo al fatto Z, di cui X considera Y responsabile". Come si nota, la forma proposizionale descrive la situazione rappresentata dal lessema, specifica la struttura di attanti semantici (indicati mediante incognite: X è il primo attante, Y il secondo attante, e così via) e permette di evitare il ricorso alla definizione per sinonimia, che spesso in lessicografia genera circoli viziosi, come sottolineato da Milićević (2006). Un lessema può avere diversi sensi, a cui corrispondono diverse definizioni, tutte riportate all'interno della parte semantica.

### 1.6.1.2 Parte sintattica

In questa parte sono specificati:

- la struttura di **attanti sintattici profondi**<sup>α</sup> del lessema;
- la corrispondenza tra questi e gli **attanti semantici**<sup>α</sup> individuati nella descrizione (una proprietà del lessema detta *diatesi*);
- le espressioni di superficie a cui gli attanti possono corrispondere (ad esempio se un attante può realizzarsi come nome, come verbo, se dev'essere introdotto da preposizione, eccetera).

La parte sintattica del lessema CRITICARE relativa alla definizione fornita sopra è complessa perché il verbo CRITICARE può dar luogo a una costruzione sintattica a tre attanti (X critica Y per Z), oppure a una costruzione a due attanti (X critica Z).

Ecco come si presenterebbe la parte sintattica del lessema CRITICARE:

	struttura a tre attanti			struttura a due attanti	
<i>Diatesi</i>	X = I	Y = II	Z = III	X = I	Z = II
<i>Realizzazione degli attanti</i>	1. N	1. N <b>obbligatorio</b>	1. <i>per</i> N 2. <i>per</i> V <sub>infin pass</sub>	1. N	1. N <b>obbligatorio</b>

Figura 13: parte sintattica della voce DEC di CRITICARE (Milićević, 2006; Mel'čuk e al., 1999)

Nella struttura a tre attanti, il primo attante sintattico (I) corrisponde al primo attante semantico (X), e deve necessariamente realizzarsi come nome (la N indicata nella seconda riga della tabella); il secondo attante sintattico (II) corrisponde al secondo attante semantico (Y), è un attante che dev'essere espresso obbligatoriamente e deve anch'esso realizzarsi come sostantivo; il terzo attante sintattico (III) corrisponde al terzo attante semantico (Z) e può realizzarsi in due modi: o come nome introdotto dalla preposizione *per* (ad esempio: criticare qualcuno *per* una decisione), o come verbo all'infinito passato, sempre introdotto dalla preposizione *per* (criticare qualcuno *per* aver deciso qualcosa).

Nella struttura a due attanti, il primo attante sintattico (I) corrisponde al primo attante semantico (X), che si realizza come nome, e il secondo attante sintattico (II) corrisponde al terzo attante semantico (Z), realizzato come nome. Un esempio è la frase "il giornale ha criticato la decisione del governo", in cui GOVERNO, che sarebbe il secondo attante semantico di CRITICARE viene espresso non come attante sintattico del verbo, bensì come attante sintattico di DECISIONE:

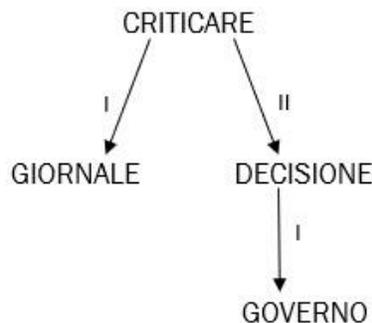


Figura 14: SSintP semplificata di "Il giornale ha criticato la decisione del governo".

### 1.6.1.3 Parte relativa alle relazioni lessicali

Questa parte raccoglie i lessemi e le espressioni con cui un dato lessema ha una relazione lessicale, ossia i suoi sinonimi, antonimi, conversivi, derivati semantici (cfr. **derivazione semantica**<sup>α</sup>) e i suoi collocati, etichettando ciascuno di essi mediante la corrispondente **funzione lessicale**<sup>α</sup>. Ad esempio, la funzione lessicale ANTI funge da etichetta per i lessemi che rappresentano degli antonimi del lessema CRITICARE; la funzione lessicale MAGN funge da etichetta per i modificatori di CRITICARE che esprimono ‘intensità, ecc. Ecco un esempio di cosa troveremmo nella parte delle relazioni lessicali di CRITICARE:

ANTI: lodare; congratularsi

MAGN: aspramente; duramente

Ovviamente i collocati possibili di CRITICARE non si limitano ai casi esposti nell’esempio: una voce del DEC contiene infatti tutte i collocati rappresentabili in termini di funzioni lessicali standard.

Con la trattazione del Dizionario esplicativo-combinatorio abbiamo concluso la presentazione della teoria Senso-Testo. Nel prossimo capitolo illustreremo l’approccio Senso-Testo allo studio delle collocazioni linguistiche e descriveremo il formalismo con cui la teoria Senso-Testo propone di rappresentarle.



## 2. Collocazioni e funzioni lessicali

### 2.1 Struttura del capitolo

In questo capitolo presenteremo il fenomeno linguistico oggetto del lavoro: le collocazioni. Dopo un'introduzione al concetto di collocazione (sezione 2.2), la cui definizione, come vedremo, è tuttora oggetto di dibattito, sarà offerta una panoramica delle principali correnti teoriche che si sono occupate dell'argomento (sezione 2.3); sarà poi approfondito il contributo dato da Igor Mel'čuk allo studio delle collocazioni, presentando la sua teoria delle funzioni lessicali come strumento formale utile a rappresentare i fenomeni collocatori (sezione 2.4). Infine, la sezione 2.5 entra nel dettaglio di alcune funzioni lessicali particolarmente significative descrivendone l'assetto semantico e sintattico.

### 2.2 Le collocazioni: introduzione

Una collocazione può essere definita come un'espressione formata da due o più unità lessicali che sono tra loro associate in modo frequente e più o meno stabile nell'uso. Pensiamo alle espressioni *pioggia torrenziale*, *beata ignoranza*, *errore marchiano*: si tratta di espressioni tipiche della lingua italiana, caratterizzate da un certo grado di fissità. Sostituendo uno dei componenti dell'espressione con un suo sinonimo otterremo un risultato percepito come innaturale: *\*sbaglio marchiano*.

Le collocazioni rivestono un ruolo fondamentale nella lingua e sono utilizzate frequentemente e naturalmente dai parlanti; lo stesso concetto di collocazione è implicitamente e inconsapevolmente posseduto da tutti: prova ne è il fatto che qualunque parlante nativo è in grado di accorgersi dell'uso imperfetto di alcune espressioni tipiche da parte di parlanti non nativi; egli direbbe che tali espressioni "suonano strane", senza tuttavia saper spiegare in termini linguistici in cosa consista esattamente l'errore, quale regola sia stata violata.

In linguistica, il ruolo delle collocazioni è stato riconosciuto solo da pochi decenni. Le collocazioni hanno da sempre suscitato l'interesse degli studiosi, ma sono state a lungo fenomeni di difficile classificazione, relegati a un ruolo marginale nella lingua, soprattutto dall'approccio generativista (Nesselhauf, 2005:1). Oggi è invece ampiamente riconosciuta

l'importanza, anche dal punto di vista numerico, delle collocazioni e in generale delle *multiword expressions* (termine di Sag e al., 2002), categoria più ampia di cui le collocazioni fanno parte:<sup>25</sup> Pawley e Syder affermano che “by far the largest part of the English speaker’s lexicon consists of complex lexical items” (1983, citato in Nesselhauf 2005:1). Per Jackendoff (1997), il lessico di un parlante nativo comprenderebbe tante *multiword expressions* quante singole parole, una stima che Sag e al. (2002:2) ritengono perfino troppo bassa, considerando che “in WordNet 1.7, for example, 41% of the entries are multiword”. La linguistica dei corpora (cfr. cap. 3 sez. 3.2.2) fornisce ulteriori prove: Howarth (1996:120), analizzando 5000 combinazioni fraseologiche del tipo nome+verbo in un corpus di 24000 parole, rileva che più di un terzo di esse sono collocazioni.

Le *multiword expressions*, e tra esse le collocazioni, svolgono una serie di importanti funzioni anche nei meccanismi di acquisizione, produzione e comprensione linguistica. Trattandosi di “blocchi preconfezionati”, diminuiscono il lavoro compositivo, e di conseguenza il carico cognitivo, da parte del locutore nella fase di produzione di enunciati,<sup>26</sup> diminuiscono il lavoro di analisi da parte del destinatario (Nesselhauf 2005:2) e, grazie alla frequenza e tipicità dell’associazione, attivano meccanismi di anticipazione che facilitano e velocizzano la comprensione (Howarth 1996:64). Inoltre, la conoscenza e l’uso corretto delle collocazioni di una data lingua sono essenziali per il raggiungimento di un alto livello di competenza della lingua stessa:

It has been recognized in English language teaching since at least the 1930s that, since native speakers employ a great many expressions that are to some degree internally fixed, someone learning a second language needs to have a command of those same expressions, to the extent that he or she wishes to approximate to native speaker proficiency. (Howarth, 1996:134)

È quest’ultimo aspetto, sottolineato da Howarth, a costituire uno dei presupposti fondamentali dello studio di caso presentato nel capitolo 4: il ruolo delle collocazioni (e le

---

<sup>25</sup> “We define multiword expressions (MWEs) very roughly as “idiosyncratic interpretations that cross word boundaries (or spaces)” (Sag e al., 2002:2). Nérima e al. (2006:95) definiscono più precisamente la categoria delle *multiword expression* come “unità lessicali costituite da più parole ortografiche non necessariamente adiacenti” (traduzione mia), evidenziando il fatto che si tratti di *unità lessicali*, e non di libere associazioni tra parole.

<sup>26</sup> “Psycholinguistic evidence indicates that the human brain is much better equipped for memorizing than for processing, and that the availability of large numbers of prefabricated units reduces the processing effort and thus makes fluent language possible.” (Nesselhauf, 2005:2)

difficoltà che esse rappresentano) nell'apprendimento di una lingua straniera. Gli altri aspetti, legati all'acquisizione linguistica e alle implicazioni cognitive delle collocazioni, non saranno approfonditi.

## **2.3 Le collocazioni: background teorico**

Le collocazioni sono state ampiamente studiate negli ultimi decenni, in particolare con l'avvento della linguistica dei corpora e del trattamento automatico delle lingue naturali. Esistono diversi approcci allo studio delle collocazioni e, di conseguenza, diverse definizioni. L'unico punto su cui gli studiosi sembrano concordare è che il termine collocazione si riferisca a un tipo di relazione sintagmatica tra parole (Nesselhauf, 2005:11) definibile in termini di co-occorrenza preferenziale (Lambrey, 2016:27), ossia che una collocazione sia un tipo di espressione formata da più parole, e che tali parole siano legate tra loro da un tipo di relazione particolare, poiché tendono ad apparire frequentemente assieme all'interno di testi.

Tra i numerosi contributi degli studiosi possiamo distinguere due approcci principali: l'approccio cosiddetto *statistico (frequency-based approach)*, le cui origini risalgono a J. R. Firth, sviluppato in particolare da Michael Halliday e John Sinclair, e l'approccio fraseologico (*phraseological approach*), fortemente influenzato dagli studi di area sovietica sulla fraseologia, e i cui principali esponenti sono A. P. Cowie, I. A. Mel'čuk e F. J. Hausmann (Nesselhauf, 2005:12).

La principale differenza tra i due approcci sta nel focus delle rispettive definizioni di collocazione. Gli esponenti dell'approccio statistico insistono sulla natura essenzialmente statistica del fenomeno, sulla frequenza della co-occorrenza delle parole che formano la collocazione, mentre i teorici che si rifanno all'approccio fraseologico trattano la collocazione come un fenomeno strutturale della lingua. Per illustrare meglio la differenza concettuale tra i due approcci dedicheremo i paragrafi che seguono ai contributi di John Sinclair e Igor' Mel'čuk.

### **2.3.1 L'approccio statistico allo studio delle collocazioni**

Un fondamentale contributo di John Sinclair allo studio delle collocazioni, è l'introduzione dell'*idiom principle*:

one of the main principles of the organization of language is that the choice of one word affects the choice of others in its vicinity. Collocation is one of the patterns of mutual choice, and idiom is another. The name given to this principle of organization is the idiom principle. The other main principle of organization which contrasts with the idiom principle is the open-choice principle (Sinclair, 1991:173).

Dunque secondo Sinclair la produzione di un testo (orale o scritto) non è il risultato di una serie di scelte che riguardano ciascuna una singola parola, bensì la scelta di alcune parole condiziona la scelta delle parole adiacenti:

a language user has available to him or her a large number of semi-constructed phrases that constitute single choices, even though they might appear to be analyzable into segments (Sinclair, 1991:110).

Le collocazioni, in quanto associazioni ricorrenti e specifiche, testimonierebbero quindi l'esistenza di un principio di distribuzione delle parole in un testo.

Sinclair definisce la collocazione come “the occurrence of two or more words within a short space of each other in a text” (Sinclair, 1991:170), intendendo per *short space* un intervallo di 4 parole a sinistra e a destra della parola presa in considerazione (ossia il *nodo*); le parole in questo intervallo sono definite *collocati* (Sinclair, 1991:170).

Sinclair opera poi una distinzione tra collocazioni significative e casuali, distinte tra loro da un criterio statistico: le collocazioni significative sono co-occorrenze di parole “such that they co-occur more often than their respective frequencies and the length of the text in which they appear would predict” (Jones e Sinclair, 1974:21, citato in Nesselhauf, 2005:12). Ad esempio, in “il sole tramonta”, “il” non è considerato un collocato significativo di “sole”, perché la sua frequenza generale è estremamente alta (si accompagna a sostantivi maschili singolari), mentre “tramonta” è un collocato significativo, perché si tratta di una parola poco frequente, che tendenzialmente si trova in contesti in cui co-occorre la parola “sole”.

La definizione di collocazione nella letteratura ispirata all'approccio statistico non è unica; diversi autori hanno diversi modi di intendere le collocazioni, e addirittura “even Sinclair sometimes varies in how he defines collocations” (Nesselhauf, 2005:13). Un elemento di variazione riguarda ad esempio l'unità di base della collocazione: la *parola* intesa come *lemma* oppure in senso più ampio? Halliday (1966:156) definisce le collocazioni come relazioni tra *lexical items*, dove per *lexical item* si intende un gruppo di parole legate morfologicamente da una relazione di derivazione; nell'esempio fornito dallo stesso Halliday

STRENGTH, STRONG e STRENGTHEN costituiscono varianti di un medesimo *lexical item*, e le collocazioni *a strong argument, to argue strongly, the strength of the argument* e *his argument was strengthened* costituirebbero varianti di una medesima collocazione. Un diverso punto di vista è quello che considera separatamente ciascuna forma di un dato lemma, analizzandone la variazione dei collocati: ciò permette di rendere conto, ad esempio, delle differenze tra i collocati di *frutto, frutti e frutta*. Un ulteriore elemento di variazione interna riguarda il numero di parole da cui può essere costituita una collocazione, e se le parole debbano essere consecutive o meno. Un'altra questione aperta riguarda la relazione sintattica tra collocato e base: i teorici che adottano l'approccio statistico solitamente non ne tengono conto, ad eccezione di Kjellmer (1994) e Greenbaum (1970), che ammettono come collocazioni solo gruppi di parole che hanno un legame grammaticale forte tra di loro (Nesselhauf, 2005:13). È possibile infine individuare altri due punti su cui l'accordo non è unanime: tali punti costituiscono due dei sei parametri adottati da Gries (2008) per la classificazione dei fenomeni fraseologici, ossia la *flessibilità* e il *criterio semantico*. La flessibilità è un parametro che indica il grado di flessibilità/fissità sintattica e lessicale della collocazione (ad esempio: una collocazione verbale può essere volta alla forma passiva? Possono essere introdotti modificatori?). Il criterio semantico riguarda invece il ruolo svolto dall'unità semantica della collocazione (cioè se la collocazione costituisce o meno un'unità di senso) e dalla sua non-composizionalità/imprevedibilità semantica (cioè se il senso della collocazione (non) sia la somma dei sensi dei suoi elementi e dunque (non) possa essere previsto). Le opinioni dei teorici dell'approccio statistico e frequency-based divergono soprattutto su questo ultimo punto.

### 2.3.2 L'approccio fraseologico e il pensiero di I. A. Mel'čuk

Il cosiddetto *approccio fraseologico* allo studio delle collocazioni, nato in seno alla linguistica russa degli anni sessanta (di cui Mel'čuk è un esponente di punta) condivide, anticipandoli, alcuni presupposti dell'*idiom principle*, in particolare l'idea che la scelta di una determinata parola da parte del locutore comporti restrizioni sulla scelta delle parole adiacenti. L'*idiom principle* è in effetti un buon punto di contatto tra l'approccio statistico e quello fraseologico, anche se i due approcci guardano a esso da due prospettive metodologiche (ed epistemologiche) molto diverse. Per Sinclair l'*idiom principle* è un principio emerso da osservazioni empiriche, che descrive il funzionamento della lingua senza ambire a darne una

giustificazione *strutturale*; Sinclair e i teorici legati al suo approccio rimangono legati al puro dato statistico e quantitativo, descrivendo le collocazioni in termini di frequenze, occorrenze, intervalli. Viceversa, Mel'čuk vede nelle restrizioni che condizionano le co-occorrenze di una parola un vero e proprio principio *strutturale*, che riflette cioè la struttura, il funzionamento stesso della lingua. Ricordiamo a questo proposito che Mel'čuk riformula la nozione Saussuriana di doppia articolazione del **segno linguistico**<sup>α</sup> aggiungendo, oltre al livello del significante e del significato, una terza articolazione: le **proprietà combinatorie**<sup>α</sup> (cfr. cap. 1 sez. 1.4). Le proprietà combinatorie di un segno linguistico sono l'insieme delle restrizioni (più o meno arbitrarie) sulle co-occorrenze del segno, ossia i diversi modi in cui esso può combinarsi ad altri segni linguistici all'interno di una data lingua naturale. Per Mel'čuk, dunque, le collocazioni riflettono una proprietà intrinseca dei segni: la matrice strutturalista del suo approccio è evidente.

La definizione generale di **collocazione**<sup>α</sup> data da Mel'čuk (1998:23) (sulla quale altri teorici come A. P. Cowie (1998) concordano) è in effetti fondata sulla nozione di restrizione: “Collocations—no matter how one understands them—are a subclass of what is known as *set phrases*”. La caratteristica fondamentale delle collocazioni è individuata dunque nella loro natura di combinazioni di parole *non libere*: questo è il significato di *set phrase*, di cui sarà data una definizione nei paragrafi successivi.

In generale l'approccio fraseologico tende a identificare le collocazioni come combinazioni di parole “a metà strada tra le *free combinations* e gli *idioms*” (Santandrea, 2014:10), ossia associazioni frequenti tra parole, ma non del tutto fisse. Mel'čuk (1998) offre una definizione formale di collocazione servendosi di una terminologia e di concetti propri della logica matematica, e sottolineando anch'egli le caratteristiche formali che distinguono le collocazioni da altri tipi di combinazioni di parole. Nel fare ciò, stabilisce una classificazione dettagliata dei diversi tipi di combinazioni di parole. Per la trattazione approfondita di tale classificazione si rimanda a Mel'čuk (1998, 2013b e 2015); in questa sede saranno citati solo i punti principali.

La prima distinzione operata da Mel'čuk è tra le *free phrases* e le *set phrases* (a queste ultime ci riferiremo, da ora in poi, con il termine **frasema**<sup>α</sup>): le prime sono combinazioni libere di parole, sottoposte unicamente alle regole sintattiche e morfologiche di una data lingua, il cui significato è uguale all'unione dei significati delle singole parole che compongono

l'espressione; i frasemi, viceversa, sono combinazioni non libere, sottoposte a restrizioni ulteriori rispetto alle regole morfosintattiche e alle regole generali della lingua: si tratta, in altre parole, di combinazioni idiosincratiche. I frasemi sono a loro volta classificati in base al tipo di restrizioni di cui sono il frutto (pragmatiche, semantiche, lessicali). Ai fini della presente esposizione, saranno presi in considerazione solo i *frasemi lessicali*.

Mel'čuk (2013b) definisce *frasema lessicale* un frasema il cui significato 'σ' viene espresso nella lingua L mediante un'associazione di parole ristretta a livello lessicale, ovvero motivata dai modi particolari in cui nella lingua L è possibile esprimere il significato 'σ'. Le collocazioni e le locuzioni idiomatiche (o *idioms*) appartengono alla categoria dei frasemi lessicali.

L'altra importante distinzione operata da Mel'čuk (2013b) è tra la nozione di **collocazione**<sup>α</sup> e quella di **locuzione**<sup>α</sup>, distinzione legata al concetto di *composizionalità*. Una locuzione è, per Mel'čuk, un *frasema lessicale non-composizionale*, cioè un'espressione fissa (non frutto di composizione) che viene selezionata "in blocco" da un locutore per esprimere un dato significato; un esempio di locuzione è *dormire sugli allori*. Una collocazione, viceversa, è un *frasema lessicale composizionale* poiché è il frutto di una composizione in cui uno degli elementi (detto *base* della collocazione) è scelto liberamente dal locutore per esprimere un senso, mentre la scelta degli altri elementi (i *collocati* della base) è condizionata dalla scelta del primo elemento, oltre che in funzione del senso da esprimere (Mel'čuk, 2013b); un esempio è l'espressione *dormire come un ghiro*, in cui la scelta dell'espressione 'come un ghiro' è determinata dalla scelta della base 'dormire', oltre che dalla volontà del locutore di esprimere il significato 'profondamente'.

La relazione che lega la base ai suoi collocati è essenzialmente lessicale. Tale **relazione lessicale**<sup>α</sup> può essere fondata su una relazione semantica esistente tra base e collocato, ma in numerosi casi la relazione semantica è assente. Citiamo a tal proposito i seguenti esempi di collocazione offerti in (Mel'čuk 2013b:7) relativi alla lingua francese (i collocati sono sottolineati, la base è riportata in maiuscolo):

porter une ACCUSATION [contre N] vs. faire ALLUSION [à N]  
dans l'ANTIQUITÉ vs. au MOYEN-ÂGE,  
salve d'APPLAUDISSEMENTS vs. vague de VIOLENCE

*Porter une accusation* (muovere un'accusa) e *faire allusion* sono costruzioni sintattiche con verbi di supporto (*light verb constructions*), in cui il verbo è semanticamente vuoto, e pertanto non può esserci relazione semantica tra base e collocato. Inoltre, la peculiarità dei verbi di supporto è l'apparente arbitrarietà con cui essi sono associati ai sostantivi: *make a choice*, *take a break*, *have a shower* vs. *fare una scelta*, *fare una pausa*, *fare una doccia*. Le dissimmetrie intralinguistiche e interlinguistiche (riscontrabili anche nell'uso delle preposizioni) sono significative per capire che il legame tra base e collocato è essenzialmente lessicale, non necessariamente motivato da altri fattori.

Mel'čuk definisce la relazione lessicale esistente tra base e collocato relazione di *co-occorrenza lessicale condizionata*.<sup>27</sup> Si parla di co-occorrenza condizionata quando due o più elementi linguistici co-occorrono (cioè compaiono assieme) non per libera associazione ma in ragione di restrizioni derivate dalle *proprietà combinatorie*<sup>α</sup> dei singoli elementi.

Possiamo proporre una definizione di collocazione che riassume il punto di vista di Mel'čuk: una collocazione è un frasema lessicale compositivo, ossia un'espressione formata da due o più parole (la base e i suoi collocati) legate da una relazione semantica o lessicale, che non sono associate liberamente, né la loro associazione è completamente fissa: una collocazione è frutto di una composizione in cui la scelta della base determina la scelta dei collocati in virtù di restrizioni lessicali, ossia di una relazione di co-occorrenza lessicale condizionata.

## 2.4 Funzioni lessicali

La sezione precedente illustra come per Mel'čuk le collocazioni siano espressioni in cui una parola (la base) è selezionata dal locutore per il suo significato, e le altre parole (i collocati) sono selezionati in funzione della base, secondo restrizioni di tipo lessicale.

Un interessante strumento di rappresentazione delle restrizioni lessicali che caratterizzano le collocazioni è offerto, all'interno del modello linguistico Senso-Testo di Mel'čuk, dalle funzioni lessicali. Numerosi studi hanno già analizzato a fondo il tema delle funzioni lessicali in relazione alle collocazioni, tra i quali citiamo Mel'čuk (1996, 1998, 2003), Wanner (1996),

---

<sup>27</sup> Termine da me coniato traducendo il termine inglese *restricted lexical co-occurrence*.

Polguère (2003), Tutin (2005), Jousse (2010). Lo scopo di questa sezione pertanto non è di fornire una trattazione esaustiva (per la quale si rimanda alle opere citate) quanto di fornire un'introduzione al tema che permetta al lettore di acquisire familiarità con il concetto di funzioni lessicali e con la loro classificazione, e di comprendere l'utilità delle funzioni lessicali ai fini del presente lavoro.

#### 2.4.1 Funzioni lessicali: definizione

Le funzioni lessicali, secondo la definizione dello stesso Mel'čuk (1996:38, traduzione mia) sono "un insieme di strumenti formali ideati per descrivere, in maniera compatta e sistematica, i tipi di relazioni lessicali esistenti tra le unità lessicali di una lingua".

La prima elaborazione teorica delle funzioni lessicali si deve a Žolkovskij e Mel'čuk (1965, 1966, 1967), il cui scopo era analizzare e rappresentare la struttura del lessico di una lingua nel quadro del nascente modello Senso-Testo, e analizzare i meccanismi operanti a livello lessicale nella costruzione di enunciati. I loro studi, come già specificato nel capitolo 1 del presente lavoro, erano orientati verso applicazioni immediate nel campo della traduzione automatica: ciò permette di comprendere l'elevato grado di formalizzazione del modello da essi proposto.

Il termine *funzione* nel contesto delle funzioni lessicali è impiegato nel suo senso matematico,<sup>28</sup> come emerge dalla definizione di funzione lessicale data da Mel'čuk (1996:39, traduzione mia):

Dal punto di vista formale, una Funzione Lessicale  $f$  è una funzione che associa a una data espressione  $L$ -l'*argomento*, o *base*, della funzione  $f$ -un insieme  $\{L_i\}$  di espressioni-il *valore* della funzione  $f$ -che esprimono, in relazione a  $L$ , un significato specifico associato a  $f$ :

$$f(L) = \{L_i\}.$$

Data la complessità della definizione, si ritiene utile offrire un esempio ripreso e adattato da Gatti (1992). Si considerino le seguenti coppie di numeri  $(x, y)$ :

---

<sup>28</sup> "In matematica [...] si dice che una quantità  $y$  (variabile dipendente) è funzione di un'altra quantità  $x$  (variabile indipendente) se esiste una legge che a ogni  $x$  fa corrispondere uno o più valori di  $y$ : tale legge prende essa stessa il nome di *funzione*, e questa corrispondenza si indica per lo più con la formula  $f(x) = y$  (nella quale la  $x$  è detta *argomento*, e la  $y$  *valore* della funzione  $f$ )" Vocabolario Treccani Online, <http://www.treccani.it/vocabolario/funzione/> [Consultato: 8/10/2017]

(3, 9)            (4, 12)            (5, 15)            (6, 18)

Ciò che accomuna tutte queste coppie di numeri è il fatto che il secondo elemento di ogni coppia (y) corrisponde al triplo del primo numero (x); la relazione che lega tutte queste coppie di numeri può essere dunque rappresentata mediante la seguente funzione:

$$3(x) = y$$

Si considerino ora le seguenti collocazioni della lingua italiana:

*stanco morto*            *povero in canna*            *ricco sfondato*

ciò che accomuna tutte queste espressioni è la relazione tra il primo elemento (la base della collocazione) e gli elementi restanti (i collocati). Tra *stanco* e *morto* esiste la medesima relazione che esiste tra *povero* e *in canna*, e tra *ricco* e *sfondato*: si tratta di una relazione di senso che possiamo genericamente definire ‘intensità’, e che a livello lessicale può realizzarsi in una grande varietà di modi, a seconda della parola che funge da base. Tale relazione, secondo Mel’čuk e Žolkovskij, può essere rappresentata a livello formale come una *funzione lessicale*, esattamente come la relazione tra numeri del primo esempio può essere rappresentata mediante una funzione matematica. Una funzione lessicale consiste dunque in una notazione formale che permette di rappresentare e descrivere relazioni lessicali esistenti tra parole di una data lingua.

Riprendendo la notazione di Mel’čuk  $f(L) = \{L_i\}$ , ed etichettando Magn la relazione di ‘intensità’, possiamo rappresentare le collocazioni dell’esempio precedente come istanze di una medesima funzione lessicale  $Magn(L) = L_i$

$Magn(stanco) = morto$   
 $Magn(povero) = in\ canna$   
 $Magn(ricco) = sfondato$

Secondo la terminologia proposta da Mel’čuk, Magn è il **nome** di questa particolare funzione lessicale (Magn è solo una tra le funzioni lessicali descritte all’interno del modello Senso-Testo), *stanco*, *povero* e *ricco* sono gli **argomenti** della funzione lessicale, e *morto*, *in canna* e *sfondato* sono i **valori** della funzione lessicale Magn.

Ciascuna funzione lessicale è identificata da un nome che è l’abbreviazione di una parola latina o greca e che funge da etichetta esplicativa: Magn, Oper, Labor, Func, eccetera (Mel’čuk, 1996:45).

A questo punto, prima di approfondire la teoria delle funzioni lessicali, è utile sottolineare un aspetto che aiuterà il lettore a comprendere la praticità delle funzioni lessicali per rappresentare i fenomeni collocatori. Nonostante le funzioni lessicali non siano state pensate primariamente come strumento di rappresentazione formale delle collocazioni, notiamo che si prestano particolarmente bene a questo scopo a causa di una corrispondenza diretta: come emerge dall'esempio della funzione lessicale Magn, l'argomento di una funzione lessicale corrisponde alla base di una collocazione, e i valori della funzione coincidono con alcuni dei possibili collocati della base stessa.

Le funzioni lessicali nascono dall'intuizione che il lessico sia permeato da rapporti costanti come quello individuato dalla funzione lessicale Magn. Le funzioni lessicali individuate da Mel'čuk consentono di rappresentare tali rapporti, di costruire una "grammatica del lessico", ossia paradigmi strutturati che, dato un lessema, indichino con quali altri lessemi esso può combinarsi per esprimere determinate relazioni lessicali costanti (Gatti, 1992:64). È importante precisare che le funzioni lessicali rappresentano relazioni essenzialmente lessicali, ma non necessariamente hanno alla base una relazione semantica: se è vero che Magn esprime il senso 'intensità', altre funzioni lessicali non hanno un senso proprio ma esprimono unicamente un ruolo sintattico: è il caso della funzione Oper, che ha come valore un verbo di supporto semanticamente vuoto (cfr. sez. 2.5.1.1).

Mel'čuk elenca 64 funzioni lessicali,<sup>29</sup> raccolte in un inventario dettagliato (Mel'čuk, 1996) e riportate in appendice al presente lavoro (Appendice 2: Funzioni Lessicali Standard Semplici). Si tratta di un inventario che è stato ed è tuttora oggetto di dibattito nella comunità accademica,<sup>30</sup> ma che abbiamo scelto di riportare nel presente lavoro perché costituisce una classificazione semplice e una buona base di partenza per avvicinarsi allo studio delle funzioni lessicali.

---

<sup>29</sup> Si tratta funzioni lessicali *standard semplici* (*simple standard lexical functions*). Nelle sezioni 2.4.2.1 e 2.4.2.3 verrà illustrato il significato di questa denominazione.

<sup>30</sup> A partire dalla pubblicazione del primo inventario di funzioni lessicali in Mel'čuk (1974), negli anni sono state proposte diverse classificazioni delle funzioni lessicali, tra cui quella di Alonso Ramos e Tutin (1996) e altre, di cui Jousse (2010) offre una rassegna critica. Sempre Jousse (2010) propone una nuova interessante classificazione delle funzioni lessicali, includendo anche alcune funzioni assenti nell'inventario di Mel'čuk. È inoltre attualmente in lavorazione un volume di Polguère e Mel'čuk che dovrebbe riportare un nuovo inventario di funzioni lessicali.

Nell'architettura del modello Senso-Testo le funzioni lessicali si collocano nel livello sintattico profondo (vedi capitolo 1 sez. 1.5.2.2): esse compaiono tra le unità lessicali profonde (cfr. *unità lessicale*<sup>α</sup>) con le quali vengono contrassegnati i nodi dell'albero sintattico profondo (Gatti, 1992:65), assieme ai lessemi e alle locuzioni (Milićević, 2006) (cfr. *lessema*<sup>α</sup> e *locuzione*<sup>α</sup>). Inoltre, esse hanno un ruolo essenziale all'interno del modello Senso-Testo poiché descrivono e rappresentano le proprietà combinatorie di ciascun lessema, un aspetto fondamentale per il modello di Mel'čuk, tanto che nel Dizionario esplicativo-combinatorio (cfr. cap.1 sez. 1.6) "per ciascuna voce lessicale si enumerano le relative funzioni lessicali, che vanno a collocarsi in una zona apposita prevista per la descrizione sistematica delle proprietà combinatorie dei lessemi" (Gatti, 1992:65).

#### **2.4.2 Classificazione delle funzioni lessicali**

Sin dalla loro prima formulazione teorica, le funzioni lessicali sono state oggetto di classificazioni di vario tipo: classificazioni basate su proprietà formali delle funzioni lessicali stesse o classificazioni più informali basate ad esempio su criteri sintattici o semantici (Mel'čuk 1974, 1996, 1998). Secondo lo stesso Mel'čuk (1998:34) "Lexical functions can be classified from different viewpoints; without having a scientific impact on the issue, such classifications facilitate the task of the user and thus possess pedagogical value".

La classificazione esposta nel presente lavoro si riferisce a quella proposta in Mel'čuk, Clas e Polguère (1995), ossia una classificazione formale tripartita: 1) funzioni lessicali standard vs. non standard; 2) funzioni lessicali sintagmatiche vs. paradigmatiche; 3) funzioni lessicali semplici vs. complesse vs. configurazioni di funzioni lessicali. Questi tre tipi di classificazione, ciascuno basato su un criterio diverso, non si escludono a vicenda, ma piuttosto si completano, evidenziando ciascuno un aspetto diverso di una medesima funzione lessicale.

##### **2.4.2.1 Funzioni lessicali standard vs. non standard**

Il primo criterio di distinzione è essenzialmente quantitativo: sono funzioni lessicali standard quelle che possono essere applicate a un elevato numero di argomenti e che hanno un elevato numero di valori; in altre parole, sono standard le funzioni lessicali che rappresentano

relazioni lessicali “generalì e universali” (Milićević, 2006)<sup>31</sup> come ad esempio la funzione lessicale Magn:

Magn(pioggia) = battente  
Magn(rumore) = assordante  
Magn(sano) = come un pesce

Viceversa, sono dette funzioni lessicali non standard quelle che rappresentano una relazione lessicale estremamente specifica, limitata a un numero ristretto di lessemi (quando non addirittura a uno solo): *anno bisestile*, *pasta al dente*, *bistecca al sangue*. Questa prima distinzione quantitativa permette di operare una importante restrizione del campo di analisi: per poter analizzare in modo sistematico i rapporti lessicali e operare generalizzazioni Mel'čuk esclude dalla propria indagine le funzioni lessicali non standard:

Non-Standard LFs cannot be logically predicted and organized into an overall system. They are numerous and extremely capricious, and have to be empirically found and registered in the corresponding lexical entries (Mel'čuk, 1996:45)

Per questo motivo esse non sono categorizzate e non hanno un nome.

È vero che le funzioni lessicali non standard permettono di rappresentare collocazioni linguistiche importanti e frequenti nell'uso, tuttavia, per ragioni di attinenza con gli scopi del presente lavoro, le funzioni lessicali non standard sono escluse anche dalla nostra indagine. Altri studiosi si sono dedicati allo studio delle funzioni lessicali non standard e ne hanno proposto una formalizzazione: si rimanda a questo proposito a Jousse (2007).

#### 2.4.2.2 Funzioni lessicali sintagmatiche vs. paradigmatiche

Un altro tipo di distinzione è quello tra funzioni lessicali sintagmatiche e paradigmatiche (Alonso Ramos e Tutin, 1996; Mel'čuk, 1996). Una funzione lessicale sintagmatica è quella che ha per argomento un lessema L, e il cui valore è un lessema L<sub>i</sub> che co-occorre con L (ossia un collocato di L). In altre parole, una funzione lessicale sintagmatica rappresenta un'espressione formata da argomento + valore della funzione. Un esempio, oltre alla già citata

---

<sup>31</sup> “A standard FL is general in that it is applicable to a high number of keywords and has many elements of its possible values. It is universal in that it is valid cross-linguistically.” (Milićević, 2006)

funzione Magn, è la funzione lessicale Oper<sub>1</sub> (il cui valore è un verbo di supporto del lessema che funge da argomento della funzione):

Oper<sub>1</sub>(dimissioni) = dare [le dimissioni]  
Oper<sub>1</sub>(pazienza) = portare [pazienza]

Una funzione lessicale paradigmatica, viceversa, ha per argomento un lessema L, e ha per valore un lessema L<sub>i</sub> che *sostituisce* L in un contesto testuale (cioè non forma con esso una collocazione):

the value f(L) of a paradigmatic LF f is typically used in the text instead of its keyword L; f(L) and L are in an "either-or" relation. (Mel'čuk, 1996:47)

Esempi tipici di funzioni lessicali paradigmatiche (tra quelle individuate da Mel'čuk) sono la funzione Syn (il cui valore è un *sinonimo* del lessema che funge da argomento della funzione), la funzione lessicale Anti (il cui valore è un *antonimo* dell'argomento della funzione), le funzioni lessicali come S<sub>0</sub>, il cui valore è un *derivato sintattico* dell'argomento (cfr. **derivazione sintattica**<sup>α</sup>), e la funzione S<sub>1</sub>, il cui valore è il *primo attante semantico*,<sup>32</sup> definito agente in altri quadri teorici, del lessema che funge da argomento:

Syn(sedia) = seggiola  
Anti(chiuso) = aperto  
S<sub>0</sub>(buono) = bontà  
S<sub>1</sub>(uccidere) = assassino

Le funzioni lessicali paradigmatiche consentono di rappresentare efficacemente alcune relazioni lessicali come la sinonimia, l'antonimia, l'iperonimia, le relazioni di derivazione (insegnare - insegnante - insegnamento, etc), e le relazioni esistenti tra i diversi attanti semantici interni a una situazione (gli attanti della situazione "omicidio" sono la vittima, l'assassino, l'arma del delitto, etc).

La distinzione tra funzioni lessicali sintagmatiche e paradigmatiche non sempre è netta, poiché esistono alcune eccezioni, e va pertanto impiegata con cautela (Lambrey, 2016:40). È possibile, ad esempio, che una funzione lessicale tradizionalmente sintagmatica come Magn ammetta come valori anche *espressioni agglutinate* (termine di Gatti, 1992:70) che di

---

<sup>32</sup> Per la definizione di attante semantico si veda la spiegazione della struttura semantica all'interno del modello Senso-Testo (cap. 1 sez. 1.5.2.1.1); cfr. **attante**<sup>α</sup> nel glossario in appendice.

fatto sostituiscono l'argomento in un contesto testuale: Mel'čuk parla in questo caso di *fusione*:<sup>33</sup>

Si parla di fusione quando il valore  $f(L)$  di una funzione lessicale  $f$  è un'unità lessicale  $L_i$  che formalmente non contiene l'argomento  $L$  di  $f$ , ma dal punto di vista semantico copre il significato di  $f$  e di  $L$ , tale che ' $L_i$ ' = ' $L + f$ '.

Le espressioni che consistono in una fusione sono convenzionalmente indicate mediante una doppia barra obliqua "/" che le separa dai valori non risultanti da fusione (Mel'čuk 1996:74, traduzione mia).

In pratica si ha fusione quando il valore di una funzione lessicale non è un'espressione che co-occorre con l'argomento della funzione, bensì un'espressione autonoma il cui senso è la somma del senso della base e del senso apportato dalla specifica funzione lessicale. Per esempio  $\text{Magn}(\text{sconfitta}) = \text{pesante} // \text{disfatta}$ . *Disfatta* è un esempio di fusione, ossia un'espressione che corrisponde a "pesante sconfitta".

La classificazione in funzioni lessicali sintagmatiche e paradigmatiche è stata studiata dai linguisti Senso-Testo fin dalla prima formulazione teorica delle funzioni lessicali, ma a causa della sua problematicità non è presa in considerazione nel presente studio.

#### 2.4.2.3 Funzioni lessicali semplici e complesse; configurazioni di funzioni lessicali

L'ultimo tipo di distinzione tra funzioni lessicali si basa sulla loro struttura (Milićević, 2006). Le funzioni lessicali semplici sono quelle che non possono essere rappresentate mediante altre funzioni lessicali; in altre parole, una funzione lessicale semplice è una funzione che rappresenta una relazione "primaria", non scomponibile (Lambrey, 2016:41). Le relazioni di questo tipo corrispondono alle 64 funzioni lessicali definite da Mel'čuk (1996). A questo proposito è opportuno ricordare che l'analisi di Mel'čuk prende in considerazione unicamente le funzioni lessicali *standard*, pertanto le 64 funzioni lessicali semplici individuate da Mel'čuk sono funzioni lessicali *standard* semplici, distinte dalle funzioni lessicali *standard* complesse. Di seguito daremo per scontato questo dettaglio e, per semplicità, ci riferiremo a tali funzioni omettendo l'aggettivo *standard*.

---

<sup>33</sup> In originale *fused element of the value of lexical functions*. Il termine italiano *fusione*, con cui propongo di identificare questo concetto, è coniato traducendo dal termine francese *fusion*, utilizzato ad esempio da Jousse (2010: 32).

Le funzioni lessicali complesse nascono dall'unione di due o più funzioni lessicali semplici. Secondo la definizione dello stesso Mel'čuk:

Una funzione lessicale complessa è una combinazione di funzioni lessicali semplici [...] il cui valore è un'espressione linguistica che ha come significato il significato della combinazione. (Mel'čuk, 1996:73, elaborazione mia)

Esempi di funzioni lessicali semplici sono le funzioni Magn e Oper<sub>1</sub>, già introdotte, mentre due esempi di funzioni lessicali complesse (formate a partire dalle prime due) sono, ad esempio, AntiMagn e IncepOper<sub>1</sub>:

**FL semplice**

Magn(ferito) = [ferito] grave

Oper<sub>1</sub>(raffreddore) = avere [il raffreddore]

**FL complessa**

AntiMagn(ferito) = [ferito] lieve

IncepOper<sub>1</sub>(raffreddore) = prendere [il raffreddore]

Nel primo esempio, la funzione Magn rappresenta una relazione lessicale semplice, dal che dal punto di vista semantico esprime 'intensità'. AntiMagn rappresenta invece una relazione semantica complessa, poiché al senso 'intensità' di Magn si aggiunge il senso 'negazione' espresso della funzione Anti, ottenendo così il senso 'non intensamente'; che in italiano è espresso, quando riferito a 'ferito', dall'aggettivo "lieve".

Allo stesso modo, nel secondo esempio, la funzione Incep (che esprime il senso 'inizio') si unisce alla funzione Oper<sub>1</sub>: il valore della funzione complessa IncepOper<sub>1</sub> è dunque un verbo di supporto il cui senso è 'iniziare a'.

Da un punto di vista formale, come specificato in Mel'čuk (1996:73), una funzione lessicale complessa è rappresentata come  $fg(L)$  e non come  $f(g(L))$ , poiché il meccanismo opera in un ordine preciso: la funzione  $f$  opera sulla funzione  $g$ , e la funzione ottenuta,  $fg$ , opera poi sull'argomento  $L$  agendo come un blocco unico.

In altre parole, il meccanismo alla base di ogni funzione lessicale complessa è il seguente: 1) le funzioni lessicali semplici di cui essa è composta operano, per così dire, ciascuna come modificatore semantico o sintattico della funzione successiva (la prima agisce come modificatore della seconda, e così via), fino a costituire un 'blocco' semantico-sintattico unitario (ossia la funzione complessa); 2) la funzione complessa così costituita 'agisce' sull'argomento  $L$ . Una funzione lessicale complessa, applicata a argomenti diversi, dà luogo a valori diversi, esattamente come se si trattasse di una funzione semplice.

Un caso a parte è rappresentato, infine, dalle configurazioni di funzioni lessicali. La loro notazione formale è  $f + g(L) = L_i$ , e la definizione fornita in (Mel'čuk, 1996:74) è la seguente (le sottolineature sono mie):

A Configuration of LFs is a combination of syntactically unrelated Simple or Complex LFs having the same keyword such that there exist a single lexical expression covering the meaning of the combination as a whole. For example:

Magn + Oper<sub>1</sub>(laughter) = roar [with ~],  
where roar means ≈ 'do [= Oper<sub>1</sub>] big [= Magn] laughter'.

In a Configuration of LFs, the syntactically central LF, which determines the part of speech of the Configuration and [of] the value, is written rightmost.

Le configurazioni di funzioni lessicali presentano alcune importanti differenze rispetto alle funzioni lessicali complesse. In primo luogo, come esplicitato nella definizione appena fornita, una configurazione di funzioni lessicali può avere tra i suoi elementi una funzione lessicale complessa.

In secondo luogo, in una funzione lessicale complessa  $fg(L)$ ,  $f$  e  $g$  agiscono come un 'blocco unico' sull'argomento  $L$ , mentre in una configurazione di funzioni lessicali  $f + g(L)$ ,  $f$  e  $g$  hanno ruoli diversi e ben definiti. Innanzi tutto, avendo lo stesso argomento (come specificato nella definizione),  $f$  e  $g$  agiscono ciascuna direttamente sull'argomento, e non l'una sull'altra;  $f$  agisce sull'argomento come modificatore semantico, proiettando il proprio senso su di esso (e non su  $g$ ), mentre  $g$  agisce primariamente come modificatore sintattico (determina cioè la parte del discorso del valore della funzione) e, secondariamente, agisce proiettando sull'argomento anche il proprio contenuto semantico (qualora  $g$  abbia un contenuto semantico).

Un esempio di configurazione di funzioni lessicali è il seguente:

Magn + Oper<sub>1</sub> (fame) = morire di fame

Le proprietà sintattiche di tale espressione derivano da Oper<sub>1</sub> (si tratta cioè di un'espressione verbale), e il suo contenuto semantico deriva da Magn (poiché Oper<sub>1</sub>, come già abbiamo visto, non ha un contenuto semantico vero e proprio). In altre parole, Oper<sub>1</sub> determina l'assetto sintattico del valore della funzione (cioè la sua parte del discorso e la sua struttura di attanti), mentre Magn "proietta" il suo senso direttamente sull'argomento e non su Oper<sub>1</sub>:

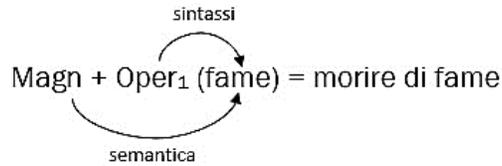


Figura 15: esempio del funzionamento di una configurazione di funzioni lessicali.

il valore di questa configurazione è infatti un'espressione il cui senso è 'avere una fame intensa' e non 'avere intensamente fame'.<sup>34</sup>

## 2.5 Approfondimenti su alcune funzioni lessicali

In questa sezione entreremo nel dettaglio di alcune funzioni lessicali, con lo scopo di mostrare come il formalismo delle funzioni lessicali permette di integrare informazioni semantiche e sintattiche in un'unica rappresentazione. Per ragioni di spazio non è possibile dedicare una trattazione approfondita a tutte le funzioni lessicali: in questa sede ci concentreremo su alcune funzioni lessicali standard semplici presenti anche nel nostro studio di caso (cfr. Cap. 4, sez. 4.4.1.1): Oper, Func, Labor, Real. Per un riassunto delle caratteristiche principali di ciascuna delle 64 funzioni lessicali di Mel'čuk si rimanda all'Appendice 2, mentre per approfondimenti ulteriori si rimanda a Mel'čuk (1996), Wanner (1996) e Jousse (2010).

Nella presente esposizione ci discosteremo leggermente dalla definizione classica delle funzioni lessicali data da Mel'čuk (1996). Nella definizione originale infatti gli indici numerici delle funzioni lessicali verbali Oper, Func, Labor, Real, ecc. si riferiscono agli *attanti sintattici profondi della base* così come definiti nella diatesi di base della parola (Mel'čuk, 1996). Nella presente esposizione invece, in accordo con il professor Lareau, si è deciso di optare per una definizione leggermente diversa, secondo la quale gli indici numerici corrispondono agli attanti semantici. Ciò si deve a due considerazioni: 1) la definizione che prende in considerazione gli attanti semantici permette di evitare di ricorrere al concetto di diatesi, e pertanto risulta più semplice per il lettore che non ha familiarità con la teoria Senso-Testo; 2) nella diatesi di base dei sostantivi si osserva una totale coincidenza tra attanti semantici e

---

<sup>34</sup> La fonte di questa spiegazione sono gli appunti del corso di Semantica tenuto da François Lareau all'Université de Montréal nell'anno accademico 2016/2017.

attanti sintattici profondi, quindi la differenza nella nostra definizione non rischia in alcun modo di distorcere la sostanza del concetto originale.

Le spiegazioni contenute nelle sezioni che seguono sono in larga parte basate su Mel'čuk (1996, 1998) e sul corso di Semantica tenuto dal professor Lareau all'Université de Montréal nell'anno accademico 2016/2017. Gli esempi presentati sono ideati da chi scrive. Le figure inserite sono state realizzate in base alle stesse convenzioni del capitolo 1: le rappresentazioni semantiche sono state semplificate omettendo i grammemi e riportando direttamente i semantemi nella forma singolare o plurale, mentre le rappresentazioni sintattiche profonde sono dotate dei grammemi essenziali (cfr. cap. 1 sez. 1.5.2.2).

### 2.5.1 Oper, Func, Labor e i verbi di supporto

Iniziamo l'esposizione presentando tre funzioni lessicali molto comuni: Oper, Func, Labor.

Queste tre funzioni sono tradizionalmente presentate assieme nella letteratura Senso-Testo perché presentano alcune importanti caratteristiche comuni: si tratta di funzioni lessicali *standard semplici* (cfr. sez. 2.4.2.1 e sez. 2.4.2.3), e sono definite funzioni lessicali *verbali*, poiché il loro valore è sempre costituito da un verbo. Oper, Func e Labor permettono di rappresentare alcuni tipi di collocazioni "verbo + sostantivo".

In particolare i valori di Oper, Func e Labor appartengono alla categoria definita *verbi di supporto* cioè *verbi semanticamente vuoti* (Mel'čuk, 1998) che co-occorrono con un sostantivo (ossia l'argomento o base della funzione lessicale) a cui fanno semplicemente da supporto verbale, senza apportare, nel contesto specifico, alcun senso proprio.

Ciò che differenzia le funzioni lessicali Oper, Func e Labor è il diverso assetto sintattico dei verbi a cui le tre funzioni lessicali corrispondono, in particolare la posizione della base nella struttura sintattica profonda (SSintP) di ciascuna funzione lessicale: nella SSintP di Oper la base è il secondo attante sintattico profondo; nella SSintP di Func la base è il primo attante, mentre nella SSintP di Labor la base è il terzo attante, come vedremo nelle sezioni successive.

#### 2.5.1.1 Oper

La funzione lessicale Oper è sempre accompagnata da un indice numerico (in cifre arabe e in pedice), che qui rappresenteremo genericamente con l'indice *i*. Il valore della funzione **Oper**,

è un verbo di supporto (semanticamente vuoto) che ha le seguenti caratteristiche a livello sintattico profondo (Mel'čuk, 1996):

- Il suo primo attante sintattico profondo (in altre parole il *soggetto*) è l'attante semantico *i* della base;
- il secondo attante sintattico profondo (il complemento) è la base stessa.

Un modo per “computare” il valore della funzione  $Oper_i$  (ossia capire quale verbo corrisponda a  $Oper_i$  in relazione a una parola-base) è il seguente: innanzi tutto occorre conoscere la struttura di attanti semantici della base (idealmente facendo riferimento a un DEC), e successivamente chiedersi “qual è il verbo di supporto che ha per soggetto l'attante *i* e per complemento la base?”.

Offriamo un esempio utilizzando come base della funzione lessicale il lessema DIRITTO, che secondo l'analisi semantica svolta da chi scrive secondo i principi della lessicologia esplicativa-combinatoria (cfr. Appendice 3: Scheda del lessema DIRITTO) possiede 4 attanti semantici:

- (1) il titolare del diritto;
- (2) l'oggetto del diritto;
- (3) l'autorità che conferisce il diritto;
- (4) gli individui o organismi “altri” di fronte ai quali questo diritto viene fatto valere, e che sono tenuti a rispettarlo.

Il valore della funzione lessicale  $Oper_1$ (DIRITTO) sarebbe il verbo *avere*, come nell'enunciato “i cittadini hanno diritto a una casa”, la cui struttura sintattica profonda è illustrata in figura 16.

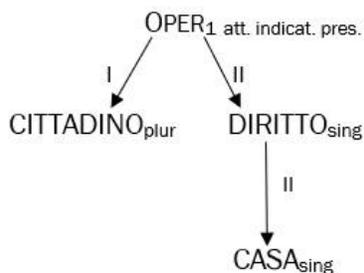


Figura 16: esempio di SSintP contenente  $Oper_1$ .

Come si nota, il verbo *avere* non è presente in questa struttura, poiché a livello sintattico profondo il nodo della struttura che spetterebbe al verbo è occupato dalla funzione lessicale corrispondente (cfr. cap. 1 sez. 1.5.2.2). Il primo attante del verbo coincide con l'attante semantico *i* del lessema DIRITTO (in questo caso il *primo* attante semantico, cioè il titolare del diritto, poiché abbiamo Oper<sub>1</sub>), mentre il secondo attante sintattico del verbo è il lessema DIRITTO.

Il verbo *avere* in questo caso è un verbo di supporto semanticamente vuoto, poiché consiste semplicemente nella realizzazione verbale del ruolo dell'attante semantico 1 di DIRITTO in lingua italiana. In altre parole, il verbo *avere* qui non apporta un senso proprio, ma è semplicemente lo strumento che la lingua italiana mette a disposizione per esprimere con una costruzione verbale il ruolo dell'attante semantico in questione. Per chiarire questo punto riportiamo la struttura semantica dell'enunciato "i cittadini hanno diritto a una casa":



Figura 17: SSem dell'enunciato "i cittadini hanno diritto a una casa".

Come emerge dalla figura 17, il verbo *avere* non compare nella struttura semantica con il suo senso di 'possesso', poiché non ce n'è bisogno: tale senso è già contenuto nel ruolo dell'attante 1.

L'indice numerico di Oper può variare: può corrispondere ad esempio al secondo o al terzo attante semantico di DIRITTO. Oper<sub>2</sub>(DIRITTO) avrebbe come valore *costituire*, come nell'enunciato "l'istruzione costituisce un diritto del bambino":

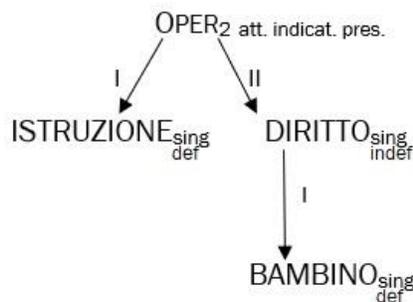


Figura 18: esempio di SSintP contenente Oper<sub>2</sub>.

In questo caso il primo attante sintattico del verbo coincide con il **secondo** attante semantico di DIRITTO (poiché abbiamo  $Oper_2$ ), cioè l'oggetto del diritto, la cosa a cui il bambino ha diritto, mentre, ancora una volta, la base della funzione lessicale costituisce il secondo attante sintattico profondo del verbo; segnaliamo inoltre che il primo attante semantico di diritto (ossia bambino) in questo caso non costituisce un attante sintattico del verbo, ma dipende dal lessema DIRITTO. Anche in questo caso, il verbo *costituire* è solo un verbo di supporto, ossia la mera realizzazione verbale del ruolo del secondo attante semantico di DIRITTO.

### 2.5.1.2 Func

Anche la funzione Func è sempre accompagnata da un indice numerico (in cifre arabe e in pedice), qui rappresentato genericamente con l'indice  $i$ . Anche la funzione lessicale  $Func_i$  corrisponde a verbi di supporto, tuttavia i verbi espressi da  $Func_i$  hanno un diverso assetto sintattico profondo rispetto ai verbi espressi da  $Oper_i$ . I verbi espressi da  $Func_i$  hanno le seguenti caratteristiche:

- il primo attante sintattico profondo (soggetto) è la base,
- l'eventuale secondo attante sintattico profondo (complemento) corrisponde all'attante semantico  $i$  della base.

Per “computare” il valore della funzione  $Func_i$  (ossia capire quale verbo corrisponda a  $Func_i$  in relazione a una parola-base) occorre dunque aver presente la struttura di attanti semantici della base, e successivamente chiedersi “qual è il verbo di supporto che ha per soggetto la base e per complemento l'attante  $i$  della base?”.

Non sempre è presente un secondo attante sintattico del verbo espresso da  $Func_i$ . È possibile infatti che la base della funzione lessicale  $Func_i$  non possieda attanti semantici, o che gli attanti della base non si trovino espressi a livello sintattico profondo come attanti del verbo. In questi casi, l'indice numerico che accompagna Func è 0. Ecco un esempio:

**Func<sub>0</sub>(PIOGGIA) = cadere (“La pioggia cade”)**

PIOGGIA non ha attanti semantici e *cadere* qui è semplicemente un verbo di supporto, ossia il verbo necessario, nella lingua italiana, a esprimere il senso ‘pioggia’ mediante una costruzione sostantivo+verbo. *Cadere* in questa costruzione non apporta un senso proprio,

corrisponde semplicemente a una realizzazione verbale del senso ‘pioggia’: infatti “la pioggia cade” e “piove” hanno lo stesso senso e la stessa struttura semantica, contrassegnata con (1) in Figura 19:

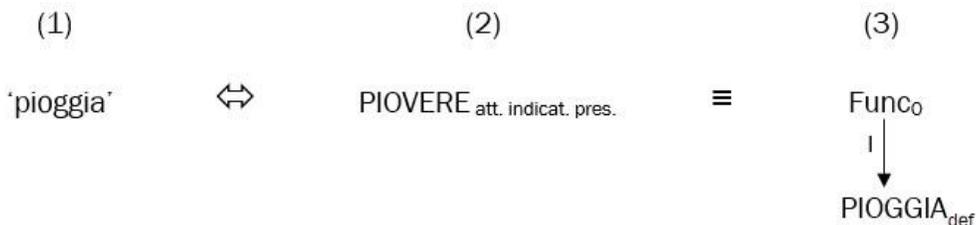


Figura 19: (1) Ssem di “piove” e “la pioggia cade”; (2) SSintP di “piove”; (3) SSintP di “la pioggia cade”.

Dal momento che il verbo *cadere* funge qui da verbo di supporto e non è portatore di un significato proprio, esso non compare nella struttura semantica. La struttura semantica (1) in figura 19 ha due possibili realizzazioni a livello sintattico profondo: la struttura (2), ossia una costruzione impersonale, e la struttura (3) in cui la funzione lessicale verbale Func<sub>0</sub> rappresenta il verbo, e PIOGGIA è il primo attante (il soggetto).

Questo tipo di costruzione con Func<sub>0</sub> è frequente con sostantivi che indicano fenomeni atmosferici, i quali tipicamente non hanno attanti.

Un altro esempio di Func<sub>0</sub> è il caso in cui gli attanti della base non siano espressi a livello sintattico profondo come attanti del verbo:

**Func<sub>0</sub>(OFFERTA) = valere** (“L’offerta vale tuttora”)

Il lessema OFFERTA ha 3 attanti semantici: (1) chi offre, (2) la cosa offerta, (3) la persona a cui si offre la cosa. Nella frase “l’offerta vale tuttora”, tuttavia, questi attanti non costituiscono attanti sintattici del verbo *valere*. Il verbo resta pertanto rappresentato dalla funzione Func<sub>0</sub>.

L’indice numerico di Func può corrispondere al primo, secondo o terzo attante semantico della base, e ovviamente al variare dell’indice varia anche il valore della funzione. Prendiamo ad esempio **Func<sub>1</sub>(DIRITTO) = spettare** (“ai bambini spetta il diritto all’istruzione”).

Poiché abbiamo la funzione Func<sub>1</sub>, è il **primo attante semantico di DIRITTO** a costituire il secondo attante sintattico del verbo (il complemento di termine, in questo caso), come è visibile anche in Figura 20.

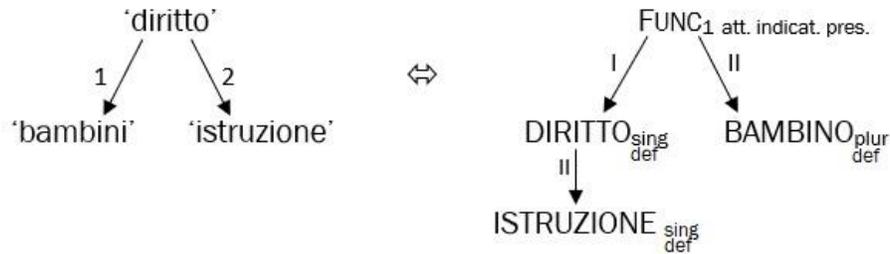


Figura 20: SSem (semplificata) e SSintP dell'enunciato "ai bambini spetta il diritto all'istruzione".

### 2.5.1.3 Labor

Labor è l'ultima funzione lessicale appartenente al gruppo dei cosiddetti verbi di supporto.

Anche la funzione Labor riporta in pedice un indice numerico. Labor ha un indice multiplo, che qui rappresentiamo con  $ij(k)$ . I verbi che sono valori di  $\text{Labor}_{ij(k)}$  hanno la seguente configurazione di attanti sintattici profondi:

- il primo attante sintattico profondo (o soggetto) è l'attante semantico  $i$  della base
- il secondo attante (o complemento) è l'attante semantico  $j$  della base
- **il terzo attante è la base**
- il quarto attante è l'eventuale attante semantico  $k$  della base.

Si spiega così il motivo per cui  $\text{Labor}_{ij(k)}$  presenta un indice numerico multiplo: la base è il *terzo* attante sintattico profondo, dunque il doppio indice numerico serve a indicare quali attanti della base andranno a costituire i restanti attanti sintattici del verbo, e a indicare l'ordine di tali attanti.

Per chiarire quanto appena esposto si riportano due esempi utilizzando come basi i lessemi INTERROGATORIO ed EREDITÀ.

**Labor<sub>12</sub>**(INTERROGATORIO)= sottoporre (“la polizia sottoporrà Battisti a un interrogatorio”)

Il lessema INTERROGATORIO ha due attanti semantici: (1) chi interroga, (2) chi viene interrogato (cfr. SSem in Figura 21). L'ordine degli indici di Labor è **1, 2**, dunque la configurazione di attanti sintattici profondi del verbo sottoporre rifletterà tale ordine, e sarà così organizzata (cfr. SSintP in Figura 21):

- il primo attante semantico di INTERROGATORIO va a costituire il primo attante sintattico del verbo

- il secondo attante semantico di INTERROGATORIO va a costituire il secondo attante sintattico
- INTERROGATORIO è il terzo attante sintattico.



Figura 21: Ssem e SSintP dell'enunciato "La polizia sottoporra Battisti a un interrogatorio".

Anche in questo caso *sottoporre* non compare nella struttura semantica, poiché trattandosi di un verbo di supporto non è portatore di un senso proprio.

Vediamo ora un altro esempio:

**Labor<sub>321</sub>(EREDITÀ)** = lasciare ("Pietro lascerà una casa in eredità ai nipoti")

il lessema EREDITÀ ha 3 attanti semantici, come il verbo ereditare: (1) chi eredita, (2) la cosa ereditata, (3) la persona da cui si eredita (cfr. Ssem in Figura 22). L'ordine degli indici di Labor è **3, 2, 1**, dunque la configurazione di attanti sintattici profondi del verbo *sottoporre* rifletterà tale ordine, e sarà così organizzata (cfr. SSintP in Figura 22):

- il terzo attante semantico di EREDITÀ va a costituire il primo attante sintattico del verbo
- il secondo attante semantico di INTERROGATORIO va a costituire il secondo attante sintattico
- **EREDITÀ è il terzo attante sintattico**
- il primo attante semantico di EREDITÀ va a costituire il quarto attante sintattico.

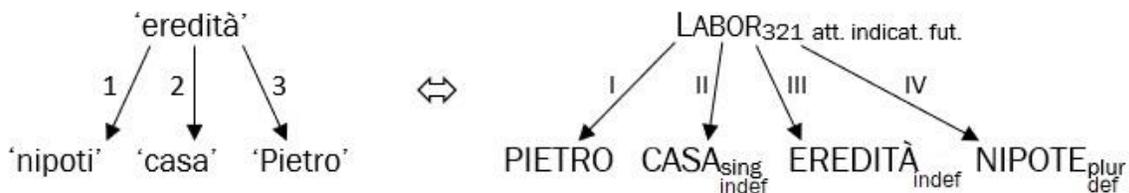


Figura 22: Ssem e SSintP di "Pietro lascerà una casa in eredità ai nipoti".

## 2.5.2 Real, Fact, Labreal e i verbi di realizzazione

Le funzioni lessicali Real, Fact e Labreal costituiscono un'altra importante famiglia di funzioni lessicali verbali. Sono funzioni lessicali standard semplici che permettono di rappresentare alcuni tipi di collocazioni "verbo + sostantivo", poiché la base di tali funzioni lessicali è sempre un sostantivo, mentre il loro valore corrisponde sempre a un verbo (Mel'čuk, 1998).

I valori di Real, Fact e Labreal appartengono alla categoria definita *verbi di realizzazione* (Mel'čuk, 1998), cioè verbi che esprimono un senso di realizzazione rispetto al sostantivo che costituisce la base della funzione lessicale. Real, Fact e Labreal ammettono infatti come base soltanto sostantivi il cui senso implica uno scopo o una funzione da realizzare, come ad esempio sostantivi che indicano artefatti (cioè oggetti concepiti per svolgere una funzione), o sostantivi astratti come il lessema PROMESSA (il cui senso implica uno scopo, poiché promettere significa 'prendere l'impegno di fare qualcosa') o il lessema DIRITTO (il cui senso implica uno scopo, cioè la tutela di una facoltà di un individuo).

Da un punto di vista sintattico Real, Fact e Labreal sono analoghi a Oper, Func e Labor rispettivamente, poiché presentano la stessa configurazione di attanti e base. In particolare, Real presenta la stessa configurazione di Oper, Fact presenta la stessa configurazione di Func, e Labor presenta la stessa configurazione di Labreal.

Ciò che differenzia le due famiglie di funzioni lessicali è semplicemente il fatto che Oper, Func e Labor corrispondono a verbi di supporto semanticamente vuoti, mentre Real, Fact e Labreal corrispondono a verbi semanticamente pieni, che esprimono appunto un senso di realizzazione.

Per chiarire questo concetto introduciamo la funzione Real. Le funzioni Fact e Labreal non saranno affrontate in questo capitolo; per la loro trattazione sintetica si rimanda alla tabella in Appendice 2.

### 2.5.2.1 Real

La funzione lessicale Real è sempre accompagnata da un indice numerico (in cifre arabe e in pedice), che qui rappresenteremo con l'indice *i*. Il valore della funzione **Real<sub>i</sub>** è un verbo di realizzazione che ha le seguenti caratteristiche a livello sintattico profondo (Mel'čuk, 1996):

- Il suo primo attante sintattico profondo (in altre parole il *soggetto*) è l'attante semantico *i* della base;
- il secondo attante sintattico profondo (il complemento) è la base stessa.

La configurazione sintattica è dunque la stessa di  $Oper_i$ . Per comprendere la differenza rispetto a  $Oper_i$  si confrontino i due esempi di seguito:

**Oper<sub>1</sub>(DIRITTO) = avere** (“il cittadino ha un diritto”)  
**Real<sub>1</sub>(DIRITTO) = esercitare** (“il cittadino esercita un diritto”)

Entrambe le funzioni lessicali hanno come base il lessema DIRITTO, il quale possiede tre attanti semantici: (cfr. sez. 2.5.1.1 e Appendice 3). La configurazione sintattica è la stessa: DIRITTO è il secondo attante sintattico, mentre CITTADINO (cioè il primo attante semantico di DIRITTO) è il primo attante sintattico (cfr. figura 23).



Figura 23: SSintP di "il cittadino ha un diritto" (a sinistra) e SSintP di "il cittadino esercita un diritto" (a destra).

Esiste tuttavia una differenza semantica importante tra i due verbi: *avere* è semplicemente un verbo di supporto, mentre *esercitare* è un verbo di realizzazione, poiché *esercitare un diritto* significa ‘realizzare lo scopo per cui un diritto esiste’, ‘fare ciò che si dovrebbe fare con un diritto’.

L'indice numerico di  $Real_i$  può variare, dando luogo a strutture sintattiche profonde in cui il soggetto del verbo di realizzazione è il secondo ( $Real_2$ ), terzo ( $Real_3$ ) o quarto attante semantico ( $Real_4$ ) della base, mentre il secondo attante sintattico è sempre la base. Di seguito riportiamo una serie di esempi utilizzando come base il lessema DIRITTO.

**Real<sub>3</sub>(DIRITTO) = garantire** (“Lo Stato garantisce il diritto alle cure mediche”)

Il terzo attante semantico di DIRITTO è l'autorità o la legge che stabilisce il diritto in questione.  $Real_3$  corrisponde pertanto a un verbo di realizzazione il cui soggetto è tale autorità (o legge).

Si tratta di un verbo di realizzazione perché esprime la realizzazione di DIRITTO ad opera dell'attante 3: nell'esempio sopra riportato lo Stato fa ciò che è suo compito fare affinché il diritto alle cure mediche si realizzi pienamente. Adottando una prospettiva leggermente diversa possiamo dire che il verbo *garantire* qui esprime la realizzazione del ruolo dell'attante 3 rispetto a DIRITTO, e dunque, di conseguenza, la realizzazione di DIRITTO.

Ecco un altro esempio:

**Real<sub>4</sub>(DIRITTO) = riconoscere** (“l'azienda ha riconosciuto il diritto dei dipendenti a un riposo settimanale adeguato”)

Il quarto attante semantico di DIRITTO è l'individuo o entità “altra” di fronte ai quali questo diritto viene fatto valere, e che sono tenuti a rispettarlo. *Real<sub>4</sub>* corrisponde pertanto a un verbo di realizzazione il cui soggetto è tale entità (in questo caso, l'azienda). Si tratta di un verbo di realizzazione perché esprime la realizzazione di DIRITTO ad opera dell'attante 4: nell'esempio sopra riportato l'azienda fa ciò che dovrebbe fare con un diritto altrui.

Spesso la funzione lessicale *Real<sub>i</sub>* corrisponde a più verbi di realizzazione, che hanno sfumature di significato diverse tra loro, o che corrispondono a diversi “gradi” di realizzazione. Ad esempio, tra i valori di *Real<sub>4</sub>* troviamo *riconoscere* e *rispettare*. *Riconoscere un diritto* significa ‘riconoscere e accettare la sua esistenza’, mentre *rispettare un diritto* corrisponde a un grado maggiore di realizzazione, poiché significa ‘comportarsi in maniera tale da consentirne l'esercizio’. Per marcare le differenze semantiche interne a una stessa funzione lessicale Mel'čuk (1996) fa ricorso a indici in apice, in questo caso indici numerici in cifre romane. Avremo così

**Real<sub>4</sub><sup>I</sup>(DIRITTO) = riconoscere**  
**Real<sub>4</sub><sup>II</sup>(DIRITTO) = rispettare**

Gli indici in cifra araba in pedice designano dunque gli attanti della base, mentre gli indici in cifre romane in apice servono a indicare diverse sfumature di senso interne a *Real<sub>4</sub>*.

Concludiamo qui la sezione di approfondimento sulle funzioni lessicali verbali. Lo scopo della sezione era mostrare come il formalismo delle funzioni lessicali permette di integrare informazioni di tipo semantico e sintattico in un'unica rappresentazione. Nel capitolo 4

vedremo come questa caratteristica del formalismo delle funzioni lessicali può trovare un'utile applicazione nell'annotazione di corpora.



## 3. Corpora e annotazione di corpora

### 3.1 Struttura del capitolo

Questo capitolo costituisce l'ultima parte di background teorico del presente lavoro. Dopo la presentazione della teoria linguistica di riferimento (capitolo 1) e del fenomeno oggetto di studio, le collocazioni (capitolo 2), introduciamo ora, sempre da un punto di vista teorico, ciò che costituirà l'oggetto dello studio di caso (capitolo 4), ossia l'annotazione di corpora. Il capitolo è diviso in due parti: nella prima (sezione 3.2) sarà offerta una definizione di corpus, una breve presentazione delle tipologie di corpora esistenti e del loro ruolo negli studi linguistici; la seconda parte del capitolo (sezione 3.3) sarà invece dedicata all'annotazione di corpora (definizione, tipologie e scopi).

### 3.2 I corpora

#### 3.2.1 Definizione

Secondo la definizione di Sinclair (1991) un corpus (pl. *corpora*) è una raccolta di testi in formato elettronico, selezionati in base a specifici criteri in modo da essere rappresentativi di una lingua o di una varietà linguistica come fonte di dati per la ricerca linguistica.

In questa definizione sono già contenuti importanti elementi per comprendere l'importanza dei corpora in linguistica. I corpora rappresentano un'importante fonte di dati linguistici, ossia la fonte di "evidenze empiriche su cui fondare lo sviluppo di modelli e teorie linguistiche" (Lenci e al., 2005). Nel corso del ventesimo secolo la linguistica ha conosciuto un'evoluzione in direzione di un crescente ricorso alla verifica empirica delle ipotesi attraverso il confronto con testi esistenti e a metodi quantitativi, basati su analisi statistiche dei dati. I corpora in linguistica svolgono dunque una duplice funzione: sono risorse per la verifica di ipotesi ma anche fonti di dati su cui formulare nuove osservazioni.

#### 3.2.2 Linguistica dei corpora

L'utilizzo dei corpora in linguistica ha iniziato a diffondersi dalla metà del ventesimo secolo grazie all'interazione di due fattori: da un lato l'applicazione di metodi empirici all'analisi

linguistica, basati sul confronto con dati testuali, e dall'altro lo sviluppo dei primi computer e la loro applicazione nell'analisi di testi (Lenci e al., 2005).

La linguistica dei corpora nasce in area anglosassone da una tradizione di ricerca linguistica di matrice empirista e si caratterizza per "l'uso di strumenti di analisi quantitativa e statistica per esplorare le regolarità linguistiche che emergono dai testi e che vengono a costituire la base per la descrizione del linguaggio" (Lenci e al., 2005:15). In questo, la linguistica dei corpora deve molto agli sviluppi dell'informatica e si intreccia strettamente con un'altra disciplina che applica strumenti informatici all'analisi linguistica, la linguistica computazionale.

Esistono pareri discordanti sullo status della linguistica dei corpora come disciplina: McEnery e Wilson (2001) la definiscono non un ramo della linguistica come possono essere la sintassi o la semantica, né una teoria, ma un approccio metodologico che può essere applicato a vari aspetti della ricerca linguistica. Altri autori come Leech (1991) lo definiscono uno specifico approccio *filosofico* allo studio della linguistica, dotato di una sua portata teorica. Tuttavia, sempre McEnery e Wilson (2001) sottolineano che la metodologia basata sull'uso di corpora può essere applicata con successo anche all'interno di ricerche linguistiche già caratterizzate da una propria impostazione teorica, e citano come esempio linguisti strutturalisti come Boas (1940) e Harris (1993) che si sono serviti di approcci *corpus-based*. Anche i linguisti che si ispirano all'approccio Senso-Testo hanno fatto ampio ricorso ai metodi della linguistica dei corpora, ad esempio Fontenelle (1992); l'oggetto stesso del presente lavoro costituisce un'integrazione tra teoria Senso-Testo e metodi della linguistica dei corpora.

Esiste però un approccio teorico incompatibile con la linguistica dei corpora, ossia l'approccio generativista, che non vede nei corpora una base affidabile su cui condurre studi. Una motivazione di questo rifiuto dei corpora da parte dei generativisti risiede nella distinzione che essi fanno tra competenza linguistica ed esecuzione, cioè "tra la conoscenza del linguaggio che hanno i parlanti e l'uso che ne fanno nelle situazioni comunicative" (Lenci e al., 2005:15). La grammatica generativa di Chomsky studia l'insieme di regole che costituiscono la competenza linguistica, e secondo Chomsky tali regole sono strutturate indipendentemente dall'uso. Pertanto le informazioni quantitative riguardanti l'uso del linguaggio (in altre parole i corpora e i dati da essi estratti) non sono rilevanti per i generativisti, e sono escluse dalle loro indagini. Nel corso degli anni, inoltre, Chomsky ha

mosso una serie di obiezioni all'uso di corpora nella ricerca linguistica. Per tali obiezioni si rimanda a Chomsky (1970) e a Andor (2004).

Il lavoro sui corpora ha offerto (e continua a offrire) importanti contributi allo sviluppo della linguistica e di discipline come la lessicologia e la lessicografia, gli studi sulla traduzione e l'interpretazione e gli studi sull'acquisizione di seconde e terze lingue (Nesselhauf, 2005). Inoltre è ormai riconosciuto il valore dei corpora come fonte di dati non solo per la ricerca, ma anche nella didattica delle lingue e nella formazione di interpreti e traduttori (Bernardini, 2004), e nella pratica stessa della traduzione e dell'interpretazione.

Tra i vari benefici dell'uso dei corpora in linguistica citiamo infine l'impulso dato allo studio delle collocazioni, fenomeno relegato dai generativisti a un ruolo marginale nella lingua (Nesselhauf, 2005), diventato oggetto di studio approfondito proprio grazie alla linguistica dei corpora e al lavoro di teorici come Michael Halliday e John Sinclair (Granger, 1998).

### 3.2.3 Criteri di valutazione dei corpora

Affinché un corpus possa rappresentare una fonte affidabile di dati deve essere conforme a criteri che esponiamo brevemente di seguito. Occorre precisare innanzi tutto che un corpus è sempre il risultato di una selezione di materiale testuale, e che questa selezione incide sui suoi possibili usi (Lenci e al., 2005). Ciò significa che uno stesso corpus potrebbe essere una buona fonte di dati per una determinata ricerca, ma inadatto agli scopi di un'altra.

Un'altra precisazione necessaria è quella sottolineata da Sinclair (2005) e riguarda la costruzione di corpora: per evitare circoli viziosi in cui gli studiosi costruiscono corpora per confermare le loro ipotesi, i criteri di selezione dei testi da inserire in un corpus dovrebbero essere sempre criteri *esterni*, ossia legati alla tipologia testuale oggetto di analisi, e non al fenomeno linguistico specifico:

Selection criteria that are derived from an examination of the communicative function of a text are called external criteria, and those that reflect details of the language of the text are called internal criteria. Corpora should be designed and constructed exclusively on external criteria (Sinclair, 2005:1).

Per ragioni analoghe, i testi contenuti nel corpus devono essere *autentici* (Sinclair, 1991), ossia testi reali preesistenti la creazione del corpus e non creati ad hoc.

Oltre a quanto esposto finora, l'affidabilità di un corpus come fonte di dati può essere valutata secondo parametri qualitativi (rappresentatività) e parametri quantitativi (dimensioni del corpus e bilanciamento).

La *rappresentatività* è un parametro qualitativo che indica quanto un corpus è in grado di essere un buon campione statistico, ossia la sua capacità di “fornirci un modello in scala” delle proprietà di una lingua o di una sua varietà” (Lenci e al, 2005:36). La rappresentatività di un corpus è strettamente legata ai parametri quantitativi di dimensione e bilanciamento.

Le *dimensioni* di un corpus sono il primo parametro da tenere in considerazione: per avere maggiori probabilità di osservare un fenomeno e poter compiere generalizzazioni affidabili occorre far ricorso a un campione di dati sufficientemente ampio. Ciò significa ad esempio che un corpus di piccole dimensioni non è adeguato a rappresentare una lingua nella sua globalità (i corpora di lingua generale come il *British National Corpus* hanno dimensioni dell'ordine delle centinaia di milioni di parole). Tuttavia, un corpus di piccole o medie dimensioni potrebbe essere adeguato a rappresentare un variante linguistica o una specifica tipologia testuale, purché esso rispetti un altro importante criterio quantitativo, il bilanciamento.

Il *bilanciamento* di un corpus indica quanto un corpus è diversificato rispetto a un ampio spettro di tipologie testuali (Biber, 1993), ossia se la proporzione tra testi appartenenti a tipologie o varietà diverse è adeguata per evitare prominenze di singole componenti che “sbilancerebbero” il corpus. Un esempio di cattivo bilanciamento di un corpus di lingua generale sarebbe dato dalla netta prevalenza di testi letterari; un esempio di cattivo bilanciamento di un corpus di italiano giuridico sarebbe la netta prevalenza di testi appartenenti all'ambito del diritto civile.

Occorre ricordare che qualunque corpus, per quanto costruito nel rispetto dei criteri sopra descritti, rimane un campione, ossia un insieme finito di materiali la cui rappresentatività, per quanto buona, avrà sempre e comunque dei limiti. Nella comunità scientifica è stata dunque riconosciuta come buona prassi fornire documentazione in cui i creatori di un corpus specificano le caratteristiche del corpus (definite secondo i parametri illustrati nella sezione 3.2.4), i criteri e gli scopi con cui è stato creato, i metodi impiegati per la campionatura dei testi e per il bilanciamento (Sinclair, 2005).

### 3.2.4 Parametri di classificazione e tipi di corpora

Dopo aver illustrato i principali parametri di valutazione di un corpus, saranno ora esposti sinteticamente alcuni parametri che permettono di classificare i corpora in una tipologia articolata. Tali parametri sono ripresi da Lenci e al. (2005).

1. **Generalità**: il grado di generalità dipende da quanto i testi che lo compongono sono stati selezionati per rappresentare la lingua nella sua globalità. In base a questo criterio è possibile distinguere tra **corpora generali** (che rappresentano la lingua in maniera completa e trasversale alle diverse varietà) e **corpora specialistici** (che rappresentano una certa varietà di lingua, oppure una singola tipologia testuale, che contengono solo testi di un dominio specifico o testi con caratteristiche particolari, come nel caso dei *Learner Corpora*, contenenti testi prodotti da individui che stanno apprendendo la lingua). I corpora generali sono spesso utilizzati come risorse di riferimento per lo studio di una lingua e per questo motivo vengono detti anche **corpora di riferimento**.
2. **Modalità**: il parametro di modalità (scritta o orale) permette di distinguere tra **corpora di lingua scritta**, **corpora di lingua orale** (contenenti *trascrizioni* di testi prodotti originariamente in forma orale), **corpora misti** (che contengono testi scritti e trascrizioni di parlato in proporzioni variabili), **corpora audio** (o *speech databases*, ossia raccolte di registrazioni audio di parlato) e **corpora multimodali** (contenenti registrazioni audiovisive di parlato, utili a descrivere anche aspetti non linguistici della comunicazione, come la gestualità). I **corpora intermodali** (Shlesinger, 2009) meritano un cenno a parte, poiché l'aggettivo *intermodale* non si riferisce alla modalità di produzione dei testi tout court, bensì alla *modalità di traduzione* che li ha generati (traduzione scritta, interpretazione simultanea, interpretazione consecutiva, ecc). I corpora intermodali sono dunque “corpora containing parallel or comparable outputs of translation and interpreting” (Bernardini e al., 2016), costituiti da testi originali e da testi che sono il risultato di diverse modalità di traduzione.
3. **Cronologia**: questo parametro rispecchia il criterio temporale con cui sono stati selezionati i testi, e permette di distinguere tra **corpora sincronici** (i cui testi sono riferiti ad uno specifico periodo, utili ad esempio per studiare una fase di sviluppo di una

lingua) e **corpora diacronici** (i cui testi appartengono a periodi diversi e consentono ad esempio di osservare l'evoluzione di una lingua).

4. **Lingua:** in base alla presenza di testi in una o più lingue i corpora possono essere distinti in **corpora monolingui, bilingui o multilingui**. Questi ultimi sono ulteriormente distinguibili in **corpora (bi- o multilingui) paralleli** e **corpora (bi- o multilingui) comparabili**. Un corpus parallelo contiene testi in una lingua L1 e testi in altre lingue che costituiscono traduzioni dei testi in L1. I corpora paralleli di solito presentano un allineamento tra testi originali e traduzioni, ossia un esplicito collegamento che unisce segmenti di testi in L1 alla loro traduzione nelle altre lingue. Un corpus comparabile “non contiene invece testi in traduzione, ma testi originali in lingue diverse” (Lenci e al., 2005:31), selezionati sulla base dei medesimi criteri, ad esempio la loro appartenenza a uno stesso genere testuale o a uno stesso dominio: è in questo senso che si intende l'aggettivo *comparabile*, perché i corpora di questo tipo permettono di mettere a confronto testi in due (o più) lingue diverse ma appartenenti a un medesimo ambito, ad esempio dei verbali di arresto, osservando come da una lingua all'altra cambiano le convenzioni testuali, la densità terminologica, ecc.; inoltre, dato che i testi sono tutti originali e non traduzioni gli uni degli altri, il confronto può essere svolto senza i condizionamenti che tipicamente derivano dall'opera di traduzione.
5. **Integrità dei testi:** alcuni corpora contengono testi integrali, altri corpora contengono invece porzioni di testi; la scelta è determinata da vari fattori, tra cui lo scopo per cui il corpus è costruito (che potrebbe essere ad esempio l'analisi dell'incipit di vari discorsi politici) o necessità legate al bilanciamento del corpus: selezionare non testi interi ma solo porzioni di lunghezza predefinita e uniforme permette di evitare che testi molto lunghi “sbilancino” il corpus (Lenci e al., 2005). D'altra parte la selezione di una porzione di testo potrebbe incidere negativamente sulla rappresentatività del corpus, poiché come sottolineato da Sinclair (2005), la porzione scelta potrebbe non essere rappresentativa delle caratteristiche del testo nel complesso. Per tutti questi motivi, la scelta di includere testi integrali o porzioni di testi deve essere adeguatamente ragionata ed esplicitata.
6. **Codifica e annotazione:** la *codifica* è la rappresentazione di dati (nel caso dei corpora, i dati sono i testi) in formato digitale. Sappiamo che un testo non è semplicemente una

sequenza di caratteri, ma è una struttura complessa e articolata su più livelli, in cui ogni livello è portatore di informazioni: in maniera analoga esistono diversi livelli di codifica di un testo, a seconda di quante e quali informazioni del testo vogliamo rappresentare in formato digitale. Il livello di codifica più elementare, detto *codifica di livello zero*, è la rappresentazione digitale (binaria) della sequenza ordinata dei caratteri del testo, senza aggiunta di informazioni di nessun tipo. Ciò che si ottiene mediante la codifica di livello zero di un testo è un file in *formato solo testo* (.txt). Un corpus può consistere dunque in un semplice file (o una serie di file) in formato .txt. Esistono poi codifiche di alto livello che arricchiscono i testi del corpus con informazioni riguardanti ad esempio la struttura dei testi, la loro suddivisione in capitoli e paragrafi, oppure informazioni meta-testuali come titolo, autore, data, ecc. In caso l'informazione codificata riguardi aspetti linguistici, come la categoria grammaticale delle parole o l'organizzazione sintattica dei testi, si parla più propriamente di *annotazione di corpora*, argomento che verrà approfondito nella seconda parte del presente capitolo (sezione 3.3). Le codifiche di alto livello possono essere realizzate mediante linguaggi di marcatura (in inglese: mark-up languages) come XML.

Tutti i corpora possiedono una codifica di livello zero, essendo per definizione raccolte di testi in formato digitale; non tutti però possiedono annotazioni: è possibile distinguere dunque tra **corpora "grezzi"** (ossia di solo testo) e **corpora annotati**. I corpora annotati si distinguono tra loro per il tipo di annotazioni di cui sono dotati, e di conseguenza per i tipi di manipolazione e analisi che è possibile realizzare su di essi tramite appositi strumenti informatici.

Ovviamente uno stesso corpus può presentare più di una caratteristica tra quelle sopra definite: i parametri di classificazione riguardano aspetti diversi che non si escludono a vicenda. Ad esempio il corpus EPTIC, sul quale è basato lo studio di caso del presente lavoro (le cui caratteristiche saranno approfondite nel capitolo 4) è un corpus specialistico, intermodale, sincronico e multilingue che funziona sia come corpus parallelo che come corpus comparabile (poiché contiene sia testi tradotti che testi originali in lingue diverse, comparabili perché appartenenti a una medesima tipologia testuale).

### 3.2.5 Strumenti per la consultazione e l'analisi di corpora

Dopo aver definito criteri di valutazione e tipologie di corpora, illustreremo sinteticamente come i corpora possono essere utilizzati. Esistono diversi metodi e strumenti informatici che consentono di consultare i corpora e analizzare i dati in essi contenuti, di cui offriremo una panoramica ispirandoci alla classificazione proposta da Kilgarrif e Kosem (2012). Tale panoramica ha anche lo scopo di presentare i software utilizzati per il presente lavoro: AntConc (Anthony, 2005) e NoSketch Engine (Rychlý, 2007).

I software più semplici, detti *concordancer*, tra i quali citiamo AntConc, offrono una serie di funzioni per esplorare i testi; si tratta di funzioni di base ma molto importanti, poiché consentono già di analizzare i dati alla luce di criteri quantitativi e qualitativi, e ricavarne interessanti osservazioni. Ecco alcune di queste funzioni:

- la compilazione delle *concordanze* di una data parola, ossia la creazione di una lista delle occorrenze della parola in un corpus, dove ciascuna occorrenza è presentata nel suo contesto di frase;
- l'analisi dei collocati, ossia la visualizzazione di quali parole co-occorrono più frequentemente con una data parola, eventualmente ordinati in base a misure di associazione statistica;
- la compilazione di liste di frequenza delle parole, in cui tutte le parole del corpus sono elencate in ordine decrescente di frequenza;
- l'individuazione di *n-grammi* frequenti, ossia di gruppi di *n* parole che compaiono spesso assieme
- l'individuazione delle *parole-chiave*, ossia parole che compaiono con una frequenza maggiore nel corpus rispetto alla loro frequenza generale, misurata mediante il confronto con corpora generali di riferimento.

Esistono poi numerose funzioni avanzate, di cui sono dotati gli strumenti informatici più complessi e progettati per applicazioni specifiche, come ad esempio Sketch Engine, uno strumento di costruzione, esplorazione e analisi di corpora pensato per l'uso in lessicografia. Sketch Engine è in grado di esplorare corpora annotati (a differenza dei concordancer più semplici come AntConc) e supportare operazioni complesse come l'estrazione e gestione automatica di terminologia, o come la compilazione di *word sketches*, ossia "il riassunto in

una pagina, compilato automaticamente a partire dal corpus, del comportamento grammaticale di una parola e dell'informazione relativa alle sue collocazioni (Kilgarrif e Kosem, 2012:44).

Uno strumento informatico di complessità minore rispetto a Sketch Engine è la sua versione open source e scaricabile, NoSketch Engine, che ha funzionalità più limitate ma offre, ad esempio, la possibilità di consultare e sfruttare le annotazioni di corpora.

Un'altra importante distinzione è quella tra strumenti informatici *computer-based*, che necessitano cioè di essere scaricati, assieme al corpus da consultare, sul computer dell'utente, e strumenti online, che permettono la consultazione di corpora tramite internet. AntConc e NoSketch Engine sono esempi del primo tipo, mentre Sketch Engine appartiene al secondo tipo.

L'ultima distinzione che scegliamo di menzionare è quella tra strumenti *corpus-related*, cioè progettati per la consultazione di uno specifico corpus (o famiglia di corpora), e strumenti *corpus-independent*, progettati per l'esplorazione di qualunque tipo di corpus. AntConc, Sketch Engine e NoSketch Engine rientrano in quest'ultima categoria; due esempi di strumenti corpus-related sono BNCweb<sup>35</sup> e XAIRA,<sup>36</sup> due interfacce progettate per la consultazione del British National Corpus<sup>37</sup> (Kilgarrif e Kosem, 2012).

### 3.3 Annotazione di corpora

#### 3.3.1 Definizione e caratteristiche generali

L'annotazione di un corpus consiste nella "codifica di informazione linguistica associata al dato testuale" (Lenci e al., 2005:211), ossia l'aggiunta di informazioni che consentono di interpretare i testi da un punto di vista linguistico (Leech, 2005) specificando ad esempio la categoria grammaticale delle parole, il loro valore all'interno della frase (soggetto, verbo principale, complemento), il senso in cui sono intese in un contesto specifico, ecc.

Un'annotazione è caratterizzata da due aspetti principali, che riguardano due piani separati:

---

<sup>35</sup> [https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/bncweb\\_top.htm](https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/bncweb_top.htm)

<sup>36</sup> <http://www.natcorp.ox.ac.uk/tools/> ; <http://xaira.sourceforge.net/>

<sup>37</sup> <http://www.natcorp.ox.ac.uk/>

- 1) quali informazioni linguistiche vengono inserite, ossia: di che tipo di annotazione si tratta (morfosintattica, sintattica, semantica...); a quale teoria linguistica si fa riferimento; la definizione delle categorie linguistiche inserite e a quali etichette corrispondono. Questo aspetto riguarda il piano strettamente linguistico ed è definito nello **schema di annotazione** (Leech, 2005);
- 2) come l'informazione viene codificata, ossia il modo in cui l'annotazione è realizzata concretamente. Questo aspetto riguarda il piano tecnologico e informatico ed è definito nello **schema di codifica** (Leech, 2005).

Schema di annotazione e schema di codifica fanno parte della documentazione che dovrebbe essere redatta a supporto dell'annotazione: come per i corpora, è considerata buona prassi fornire documentazione che illustri criteri e metodi di costruzione di un'annotazione (Leech, 2005).

Come è realizzata concretamente un'annotazione? Esistono vari modi in cui le informazioni possono essere rappresentate in formato digitale e aggiunte al testo.

Un corpus, come sappiamo, consiste in una raccolta di testi in formato digitale, generalmente codificati in formato solo testo (.txt). Un modo semplice per realizzare un'annotazione che indichi la categoria grammaticale delle parole (un tipo di annotazione detta *part of speech* (*pos*) *tagging*, cfr. sez. 3.3.3.1), potrebbe essere quella suggerita da Leech (2005), ossia l'aggiunta di un'etichetta dopo ciascuna parola, e di un trattino basso che lega ciascuna parola alla propria etichetta, come nell'esempio seguente:

```
Paolo_NP regala_V un_ART libro_NN a_PREP Marina_NP1
```

L'etichetta NP identifica i nomi propri, NN i nomi comuni, V i verbi, ART gli articoli, PREP le preposizioni.

Un'annotazione di questo tipo verrebbe dunque aggiunta direttamente al testo, e codificata assieme a esso in formato .txt, in maniera piuttosto diretta e semplice; in questo modo però l'annotazione risulterebbe inseparabile dal corpo del testo, con conseguenti problemi:

- la leggibilità del testo è compromessa;
- il corpus diventa più difficile da esplorare, poiché l'annotazione costituisce "rumore" che non consente di osservare le parole nel loro contesto "naturale";

- qualora si aggiungessero annotazioni di altro livello, il testo risulterebbe ulteriormente appesantito, e sarebbe complicato separare le annotazioni l'una dall'altra
- un'annotazione non sempre è utile agli scopi di una ricerca, dunque dovrebbe essere garantita la possibilità di consultare il testo nella sua forma originale non annotata.

Per questi e altri motivi la separazione tra annotazione e corpo del testo è stata riconosciuta come principio e buona prassi: “Any information about a text other than the alphanumeric string of its words and punctuation should be stored separately from the plain text and merged when required in applications” (Sinclair, 2005).

Le annotazioni in cui dato testuale e informazione linguistica sono separati sono dette *annotazioni stand-off* e presentano numerosi vantaggi (che citiamo da Lenci e al., 2005:234):

- il dato testuale primario si mantiene leggibile e disponibile nella sua forma originale, e non viene alterato dal processo di annotazione;
- è possibile realizzare annotazioni diverse e su diversi livelli, e tali annotazioni si mantengono separate le une dalle altre; in questo modo è possibile realizzare anche annotazioni “incompatibili” tra loro (cioè annotazioni che strutturano i dati secondo categorie e differenti);
- si creano i presupposti per ricerche linguistiche basate sull'interazione di informazioni contenute in diversi livelli di annotazione (ad esempio ricerche sull'interfaccia semantica-sintassi).

Un corpus dotato di una o più annotazioni stand-off

si presenta come un ipertesto organizzato in una serie di moduli: il *modulo base*, che includerà il dato testuale primario [...] e un *modulo extra* indipendente [...] per ogni livello di annotazione linguistica. Ogni modulo extra conterrà le informazioni relative all'annotazione vera e propria, più i dispositivi di puntamento necessari per ancorare l'annotazione alla porzione di testo rilevante” (Lenci e al., 2005:235).

Un modo per realizzare corpora con annotazioni stand-off è la codifica di testi e annotazioni mediante il linguaggio XML, che attualmente costituisce uno standard internazionale, associato ad altri linguaggi che consentono di realizzare i necessari collegamenti ipertestuali (Lenci e al., 2005).

### 3.3.2 Scopo dell'annotazione

Lo scopo generale di un'annotazione di corpora è rendere esplicita, interpretabile ed esplorabile dal computer la struttura linguistica implicita nel testo (Lenci e al., 2005). Esistono poi scopi specifici per ciascun tipo di annotazione, che saranno approfonditi nella sezione 3.3.3, quando i vari livelli di annotazione saranno presi in esame.

In generale possiamo affermare che la maggiore utilità dell'annotazione risiede nell'aumento del contenuto informativo del corpus, poiché grazie all'annotazione i dati testuali “puri” vengono arricchiti di significato e inquadrati in uno schema di rapporti e categorie, in modo tale che da semplici *dati* essi diventano *informazioni* (Lenci e al., 2005:64) che possono essere utilizzate per scopi di studio e ricerca.

Inoltre, un corpus annotato consente di effettuare analisi linguistiche automatiche molto complesse, rendendo di fatto possibile la manipolazione del testo da parte del computer. L'analisi automatica di corpora annotati può essere sfruttata per scopi di studio e ricerca, per la realizzazione di dizionari, e in applicazioni ben più sofisticate. I corpora annotati sono infatti utilizzati anche in linguistica computazionale, dove trovano diverse applicazioni nel *machine learning*, nella realizzazione di motori di ricerca, di sistemi di traduzione automatica e di sistemi di sintesi e riconoscimento vocale (Jurafsky e Martin, 2008; Clark e al., 2010).

### 3.3.3 Livelli di annotazione

Sappiamo che l'annotazione consiste nella codifica di informazioni linguistiche associate al dato testuale di un corpus. Sappiamo anche che l'informazione linguistica di cui un testo è portatore è complessa, e può essere analizzata su diversi livelli. Tipicamente, l'annotazione di corpora avviene in relazione ai tradizionali livelli di descrizione linguistica: morfologia, sintassi, semantica, ecc (Lenci e al., 2005); in altre parole esiste un livello di annotazione per ogni livello di descrizione linguistica. Di seguito illustreremo i principali livelli di annotazione di corpora basandoci sull'esposizione di McEnery e Wilson (2001), Leech (2005) e Lenci e al. (2005).

#### 3.3.3.1 Annotazione per parti del discorso

L'annotazione per parti del discorso (detta anche *pos tagging* o *annotazione morfosintattica*): consiste nell'aggiunta di etichette (in inglese *tags*) che indicano la categoria grammaticale (o

parte del discorso, dall'inglese *part of speech*) di ogni parola o token,<sup>38</sup> con eventuale indicazione di informazioni morfosintattiche come genere e numero di aggettivi e nomi, persona, modo e tempo per i verbi, ecc. (Lenci e al., 2005). Il pos tagging rappresenta un livello di annotazione essenziale, di cui sono dotati tutti i maggiori corpora attualmente esistenti.

Grazie a questo tipo di annotazione è possibile esplorare un corpus e condurre ricerche mirate senza problemi legati all'omografia: ad esempio è possibile cercare tutte le occorrenze del nome maschile *danno* distinguendolo dalle occorrenze della forma verbale *danno*. Per quanto possa sembrare banale, la possibilità di disambiguare gli omografi è una condizione indispensabile senza la quale l'esplorazione di corpora diventa un compito assai difficoltoso. Questo tipo di disambiguazione è molto utile anche in linguistica computazionale, poiché il trattamento automatico di un corpus risulta assai difficoltoso se il corpus non è dotato di pos tagging.

Questo tipo di annotazione è inoltre essenziale per realizzare i livelli di annotazione successiva, ad esempio l'annotazione sintattica e la lemmatizzazione: l'individuazione della categoria grammaticale delle parole è fondamentale per determinarne la funzione sintattica e ricondurle alla loro forma base.

### 3.3.3.2 Lemmatizzazione

La lemmatizzazione consiste nell'aggiunta di etichette che indichino il *lemma* a cui ciascuna parola fa riferimento. Il lemma corrisponde approssimativamente alla forma base delle parole, cioè quella tipicamente riportata in un dizionario: l'infinito per i verbi, il maschile singolare per gli aggettivi, ecc. (Lenci e al., 2005). Ad esempio *andrete, vai, vanno* e *andassero* sono forme flesse di un unico lemma (*andare*), pertanto ciascuna di queste forme sarà dotata di un'etichetta recante la forma *andare*.

---

<sup>38</sup> Un *token* è l'unità di base dell'analisi computazionale del testo (Lenci e al., 2005). Un testo in formato digitale è "visto" dal computer come un flusso ininterrotto di codice binario; per far sì che il computer "veda" un testo come una serie di unità distinte (i token), è necessaria un'operazione di segmentazione del testo. Questa operazione di segmentazione è detta tokenizzazione. Il risultato di tale operazione è un file di testo in cui ogni token occupa una riga (Lenci e al., 2005).

Al pari del pos tagging, la lemmatizzazione facilita le ricerche mirate ed è un prerequisito fondamentale per l'analisi automatica di corpora: ad esempio, grazie alla lemmatizzazione è possibile estrarre ed esaminare ogni variante di un dato lemma senza dover ricercare ogni variante separatamente (McEnery e Wilson, 2001).

### 3.3.3.3 Annotazione sintattica

L'annotazione sintattica (detta anche **parsing**) consiste nella rappresentazione della struttura sintattica dei testi del corpus; una volta individuate le categorie morfosintattiche delle parole attraverso il pos tagging è infatti possibile identificare (e annotare) le relazioni sintattiche tra di esse (McEnery e Wilson, 2001).

Il modo in cui questo tipo di annotazione è realizzata dipende dalla teoria linguistica usata come riferimento: esistono diversi approcci teorici all'analisi sintattica, che utilizzano categorie diverse per rappresentare la struttura di una frase (Lenci e al., 2005). Gli approcci all'analisi sintattica sono sostanzialmente due:

- **L'analisi a costituenti** associa ad ogni frase una struttura gerarchica ad albero, in cui le parole sono raggruppate in costituenti (sintagmi) di dimensioni maggiori, che a loro volta sono incassati dentro costituenti di ordine superiore, costituendo così una gerarchia. Questo tipo di analisi è tipica della tradizione generativista, della grammatica trasformazionale e dei modelli minimalisti;
- **L'analisi a dipendenze** associa ad una frase una struttura gerarchica (detta *grafo aciclico orientato*), in cui le parole sono legate tra loro da relazioni binarie di dipendenza; un esempio di relazione di dipendenza è quella che lega un soggetto al suo verbo, o quella che lega un aggettivo al nome a cui è riferito. Diverse teorie adottano questo approccio, tra cui ad esempio la teoria Senso-Testo, la cui descrizione del livello sintattico della lingua si basa su strutture a dipendenze (cfr. cap. 1, sez. 1.5.2.2).

A seconda dell'approccio teorico adottato avremo dunque un'annotazione caratterizzata da un tipo di rappresentazione sintattica diversa, dotata di categorie diverse e in cui informazioni diverse vengono esplicitate e annotate.

Un corpus annotato sintatticamente è detto anche *treebank* (McEnery e Wilson, 2001), un termine che fa riferimento agli alberi che tradizionalmente sono usati in sintassi per rappresentare le strutture di frasi.

L'annotazione sintattica di un corpus permette di realizzare operazioni complesse come l'analisi dei contesti sintattici più frequenti di una data parola (ad esempio se un verbo è utilizzato più frequentemente in costruzioni transitive o intransitive). Inoltre i corpora annotati possono essere utilizzati in linguistica computazionale come base di dati per addestrare sistemi di traduzione automatica o analizzatori sintattici automatici (detti *parser*).

#### 3.3.3.4 Annotazione semantica

L'annotazione semantica consiste nell'esplicitazione del significato o contenuto semantico delle espressioni linguistiche di un testo (Lenci e al., 2005). Questa annotazione può assumere varie forme, in base alla teoria linguistica usata come riferimento e alle categorie che essa definisce. Di seguito presentiamo due tipologie di annotazione semantica.

Un possibile tipo di annotazione semantica consiste nell'etichettatura delle *parole lessicali* del corpus (escludendo le *parole grammaticali* come articoli, congiunzioni ecc.) in base a categorie semantico-concettuali che consentano di definire il significato di tali parole (Lenci e al., 2005). La granularità delle categorie può essere variabile, si possono avere categorie molto generali (come *persona, luogo, evento, oggetto, azione, risultato*) oppure categorie e sotto-categorie specifiche, definite in base a risorse lessicali come WordNet (Fellbaum, 1998).

L'utilità di questo tipo di annotazione consiste nella possibilità di condurre ricerche e analisi mirate all'interno del corpus, ad esempio un'analisi delle occorrenze e dei collocati della parola *capo* intesa come "persona responsabile", scartando tutte le occorrenze di *capo* nel senso di "testa" o "estremità". Inoltre, un corpus così annotato può essere utilizzato per "studiare la distribuzione dei sensi di una parola in relazione a specifici contesti sintattici" (Lenci e al., 2005), oppure può essere utilizzato come fonte di dati per costruire sistemi computazionali di disambiguazione, utili in applicazioni come la traduzione automatica.

Un altro tipo di annotazione semantica (compatibile e combinabile con la precedente) consiste nell'etichettatura delle parole lessicali in base al *ruolo semantico* che esse svolgono in un determinato contesto o situazione (Lenci e al., 2005), dove per ruolo semantico possono

intendersi le relazioni tematiche definite dalla grammatica generativa (agente, tema, destinatario ecc.) oppure gli attanti semantici del modello Senso-Testo (cfr. cap.1, sez. 1.5.2.1.1). Ad esempio, un'annotazione semantica di questo tipo per la frase "Paolo regala un libro a Marina" potrebbe assegnare a *Paolo* l'etichetta corrispondente al primo attante semantico del predicato *regalare*, mentre *libro* sarebbe etichettato come secondo attante e *Marina* come terzo attante; quest'annotazione rifletterebbe dunque la struttura semantica della frase così come definita nel modello Senso-Testo. Chiaramente le categorie (e le etichette corrispondenti) variano in base alla teoria linguistica utilizzata come riferimento.

L'annotazione per ruoli semantici è utile in quanto permette di estrarre dai corpora informazioni relative alla struttura di argomenti (o attanti semantici) dei predicati, e sui vincoli che tali argomenti (o attanti) devono rispettare (Lenci e al., 2005). Ad esempio analizzando un corpus annotato per ruoli semantici osserveremmo che *mangiare* è un predicato che ammette come primo attante solo soggetti animati (a meno che il verbo non sia inteso in senso figurato); questo è un esempio banale, ma è possibile applicare lo stesso metodo ad analisi più complesse e dall'esito meno scontato, ottenendo informazioni difficilmente accessibili mediante la sola intuizione.

### 3.3.3.5 Altri tipi di annotazione

Oltre ai tipi sopra elencati esistono vari altri tipi di annotazioni linguistiche. L'*annotazione fonetica* ad esempio consiste nell'aggiunta di informazioni riguardanti la pronuncia delle parole di un corpus di lingua orale (Leech, 2005), ed è particolarmente utile in linguistica computazionale per addestrare di sistemi di riconoscimento automatico di lingua parlata e sintesi vocale (Jurafsky e Martin, 2008; Tamburini, 2005). Esiste inoltre l'*annotazione prosodica*, che consiste nell'aggiunta, in un corpus di lingua orale, di informazioni riguardanti aspetti prosodici come l'intonazione, il ritmo, le pause (Leech, 2005). L'*annotazione pragmatica* consiste invece nell'aggiunta di "informazioni riguardanti il tipo di atto linguistico (o dialogico) che avviene nel contesto di un dialogo: ad esempio l'espressione OK in contesti diversi può rappresentare una conferma, una richiesta di feedback, un assenso o un segnale pragmatico che indica l'inizio di una nuova fase della discussione" (Leech, 2005:18). Un altro tipo di annotazione linguistica è l'*annotazione discorsiva*, che consiste nell'aggiunta di informazioni relative al livello dell'organizzazione del discorso, ad esempio evidenziando ed

esplicitando le *relazioni anaforiche* esistenti in un testo, utili per lo studio dei meccanismi di mantenimento e gestione della coerenza testuale (Lenci e al., 2005).

Esistono infine altri tipi di annotazione che non corrispondono, come nel caso delle annotazioni sopra elencate, a uno specifico livello di analisi linguistica. Un corpus infatti può essere analizzato da altri punti di vista, e le informazioni di cui esso è portatore possono essere esplicitate mediante annotazioni ad hoc. Un esempio tratto da Leech (2005) riguarda l'annotazione dei cosiddetti *learner corpora*, ossia raccolte di testi prodotti dai discenti di una lingua (Nesselhauf, 2005). Tali corpora sono utilizzati per studiare le caratteristiche della produzione linguistica di parlanti non nativi, e potrebbero essere annotati mediante "error tags", ossia etichette che specifichino dove sono stati commessi errori, di che tipo di errori si tratta, ecc. (Leech, 2005); questo tipo di annotazione rende più agevole l'analisi automatica degli errori. Un altro esempio è l'annotazione oggetto dello studio di caso descritto nel capitolo 4. Si tratta di un tipo di annotazione diversa dalle annotazioni linguistiche "classiche", poiché non riguarda un livello linguistico in particolare, bensì un singolo fenomeno: le collocazioni, annotate secondo uno schema ispirato alle funzioni lessicali di Igor Mel'čuk (cfr. cap. 0).

A chiusura del capitolo proponiamo una riflessione tratta da Leech (2005) e Sinclair (2005). Un corpus costruito secondo le buone prassi indicate in Sinclair (2005), rappresentativo e ben bilanciato, contiene *dati autentici* e, per così dire, *oggettivi*, che permettono generalizzazioni affidabili. L'annotazione di corpora è un'operazione più delicata, perché se è vero che permette una più agevole esplorazione e del corpus e un suo utilizzo in applicazioni computazionali, essa implica sempre un'interpretazione dei dati autentici alla luce di un preciso riferimento teorico o metodologico; in altre parole, l'annotazione non è mai un'operazione neutrale, poiché le categorie su cui essa si basa non sono oggettive e universali, ma dipendono dell'approccio teorico scelto dall'autore dell'annotazione (Leech, 2005, Sinclair, 2005). Per questo motivo quando si realizza un'annotazione di corpus occorre sempre specificare, nello schema di annotazione (cfr. sez. 3.3.1), quale approccio teorico è stato adottato come riferimento (Leech, 2005).



## 4. Studio di caso: uso delle funzioni lessicali nell'annotazione di corpora

### 4.1 Struttura del capitolo

Il presente capitolo costituisce uno studio di caso il cui scopo è presentare una possibile applicazione delle funzioni lessicali nel trattamento delle collocazioni. Il capitolo è diviso in quattro parti: dopo aver presentato l'idea alla base dello studio di caso (sezione 4.2), nella sezione 4.3 sarà offerta una breve presentazione del corpus EPTIC - *European Parliament Translation and Interpreting Corpus* (Bernardini et al. 2016) e dei criteri con cui sono state selezionate le porzioni di testi da annotare. La sezione 4.4 presenta l'annotazione per funzioni lessicali: caratteristiche generali, schema di annotazione e di codifica, modalità di consultazione dell'annotazione attraverso il software NoSketch Engine. Infine, la sezione 4.5 presenta le conclusioni sullo studio di caso.

### 4.2 Introduzione allo studio di caso

Il presente studio di caso ha come oggetto una proposta metodologica per integrare informazioni sul fenomeno delle collocazioni all'interno di un corpus, in modo da renderle consultabili e fruibili come strumento per l'apprendimento delle collocazioni, destinato in particolare agli studenti di traduzione e interpretazione non italofoni. Tale proposta consiste in un'annotazione per funzioni lessicali del corpus EPTIC.

L'idea alla base della proposta nasce dalla consapevolezza dell'importanza delle collocazioni e dalla volontà di ideare uno strumento utile al loro apprendimento. Le riflessioni da cui tale idea è nata saranno oggetto delle prossime sezioni.

Ciò che andiamo a esporre costituisce una riflessione di chi scrive, nata da un confronto tra alcune delle principali risorse disponibili per lo studio delle collocazioni, alcune riflessioni presenti in letteratura e la propria esperienza di studente di traduzione. Non si intende presentare un'analisi approfondita ed esaustiva delle risorse (lessicografiche e non) disponibili, che meriterebbe un discorso più ampio e un lavoro di tesi dedicato. In questa sede ci limiteremo a presentare alcune idee, supportate dalle riflessioni di studiosi che si sono

occupati del tema, al fine di individuare i principali limiti di tali risorse, e proporre un metodo originale che potrebbe consentire di superare tali limiti.

#### **4.2.1 Le collocazioni nell'apprendimento linguistico**

Nell'introduzione al capitolo 2 abbiamo illustrato l'importanza delle collocazioni nella struttura della lingua, nell'apprendimento e nella produzione linguistica (cfr. sez. 2.2). Negli ultimi decenni numerosi autori hanno studiato il rapporto tra collocazioni e acquisizione di una seconda lingua (tra cui citiamo in particolare Howarth, 1996, 1998 e Granger, 1998; per una rassegna completa si rimanda a Nesselhauf, 2005), ed è ormai ampiamente riconosciuto nella letteratura che l'uso corretto delle collocazioni rappresenta, da una parte, una componente essenziale per il raggiungimento di un'elevata competenza linguistica, ma dall'altra parte costituisce un elemento di difficoltà anche per gli studenti di livello avanzato. Nella formazione di interpreti e traduttori questo aspetto assume un'importanza ancora più rilevante, poiché una buona traduzione (o interpretazione) dovrebbe, in teoria, raggiungere il livello di qualità di un testo prodotto da un madrelingua e di conseguenza dimostrare una perfetta padronanza delle collocazioni da parte del traduttore (Primož, 2011:132). Per tutti questi motivi negli ultimi vent'anni si sono moltiplicati gli sforzi per inserire la componente fraseologica e relativa alle collocazioni nell'insegnamento delle lingue, attraverso l'elaborazione di nuovi metodi e approcci didattici incentrati proprio su questi aspetti (Nattinger e DeCarrico, 1992; Nesselhauf, 2005). Parallelamente, anche la lessicografia ha dedicato una maggiore attenzione alle associazioni frequenti e tipiche tra parole (Tutin, 2005): oltre alla creazione di specifiche risorse lessicografiche come i dizionari di collocazioni, possiamo notare che anche i dizionari "tradizionali" dedicano ora più spazio alle parole che co-occorrono con un dato lemma (Primož, 2011). Si veda a tal proposito il *Nuovo vocabolario di base della lingua italiana* di Tullio De Mauro (De Mauro, 2016), dotato di un'apposita sezione dedicata alle co-occorrenze tipiche di ogni voce.

Tutto questo sembrerebbe suggerire che esista un'ampia scelta di materiali utili all'apprendimento delle collocazioni e che, proprio in ragione della crescente attenzione dedicata a questo fenomeno, lo studio delle collocazioni oggi sia più agevole di quanto lo era in passato, quando erano escluse dall'indagine linguistica e dalla didattica. Eppure, specialmente nella formazione di traduttori e interpreti, l'apprendimento delle collocazioni

resta un punto critico (Primož, 2011), e osserviamo che in realtà l'apparente abbondanza di risorse non ha risolto alcune difficoltà di fondo, perché le risorse disponibili presentano alcuni limiti importanti.

#### 4.2.2 Risorse per l'apprendimento delle collocazioni: i dizionari specializzati

Come osservato da Primož (2011), non esistono molti dizionari dedicati specificamente alle collocazioni. Limitando le nostre considerazioni alla lingua italiana, che sarà oggetto del nostro studio di caso, rileviamo che attualmente i due soli dizionari di collocazioni disponibili sul mercato sono il *Dizionario delle combinazioni lessicali* di Urzì (2009) e il *Dizionario delle collocazioni* di Tiberii (2012).

A quest'ultimo può essere mossa una critica che già più di vent'anni fa Smadja (1991) e Fontenelle (1992, 1998) avevano rivolto ai dizionari di collocazioni di lingua inglese, ossia che i collocati vengono raggruppati semplicemente in base alla loro parte del discorso, senza indicazioni semantiche. Nel dizionario di Tiberii (2012), infatti, i collocati sono divisi in sezioni in base alla loro parte del discorso, e in ciascuna sezione sono elencati in ordine alfabetico, senza alcuna distinzione di senso. Ciò può risultare particolarmente frustrante per un utente non italofono, a cui Tiberii non fornisce gli strumenti per scegliere il collocato più corretto per esprimere un determinato concetto.

Anche il dizionario di Urzì (2009) presenta una macro-divisione dei collocati in base alla parte del discorso ma, diversamente da Tiberii, in ogni sezione Urzì divide i collocati per aree semantiche (ad esempio i collocati verbali dotati di sensi simili sono raggruppati e divisi, mediante barre verticali, dai collocati verbali appartenenti a un'area semantica diversa. Cfr. Figura 24); tuttavia tali aree semantiche non sono etichettate esplicitamente, né sono indicate le differenze tra sfumature interne a un'area. Rispetto a Tiberii, Urzì fornisce maggiori indicazioni: ad esempio segnala, mediante il simbolo ⇔, i collocati che costituiscono antonimi l'uno dell'altro; offre inoltre alcuni esempi d'uso dotati di valore esplicativo (cfr. Figura 24, dove il collocato è indicato in grassetto all'interno dell'esempio l'esempio “la novazione è il contratto con il quale debitore e creditore stabiliscono di **estinguere** un'obbligazione”); tali esempi tuttavia sono distribuiti in maniera non sistematica nel dizionario.

#### OBBLIGATO

**moralmente** ⇔ **legalmente, giuridicamente, formalmente; tassativamente** obbligato a  
...

#### OBBLIGAZIONE

V<sub>1</sub> (*vincolo giuridico*) **assumere, contrarre** un'obbligazione contrattuale | **assolvere, adempiere** un'obbligazione; il valore liberatorio della moneta corrisponde al suo valore nominale, per cui una banconota da 100 mila lire può essere utilizzata per **soddisfare** un'obbligazione di pari importo; la novazione è il contratto con il quale debitore e creditore stabiliscono di **estinguere** un'obbligazione sostituendola con un'altra | **liberare, sciogliere, svincolare** da un'obbligazione

(*titolo di credito*) **emettere** obbligazioni | **collocare** obbligazioni presso il pubblico | **sottoscrivere; negoziare** obbligazioni

A (*vincolo giuridico*) || (dir.) obbligazione **solidale; obbligazione parziaria**  
(*titolo di credito*) || obbligazioni **pubbliche, private; obbligazioni convertibili; obbligazioni indicizzate; obbligazioni nominative, obbligazioni al portatore**

N (*titolo di credito*) emettere una **tranche** di obbligazioni

Figura 24: esempio di voci del dizionario delle combinazioni lessicali (Urzi, 2009: 539).

Riteniamo il che dizionario di Urzi, più ricco di informazioni rispetto al dizionario di Tiberii, costituisca un ottimo strumento per utenti italo-foni (che costituiscono i destinatari primari dell'opera, per stessa ammissione dell'autore, cfr. Urzi, 2009:5). Tuttavia l'assenza di indicazioni semantiche esplicite potrebbe rendere difficoltoso il suo uso da parte di non italo-foni, poiché si rende necessario il ricorso ad altre risorse, ad esempio dizionari di lingua generale in cui andare a cercare il significato dei vari collocati.

In sintesi ci sentiamo di mettere in discussione l'utilità di questo tipo di dizionari per l'apprendimento delle collocazioni dell'italiano da parte di non italo-foni. Pur costituendo database estremamente ricchi, i dizionari di collocazioni attuali non aiutano lo studente straniero a orientarsi tra di esse, poiché "deve trovare da sé la natura precisa del legame semantico tra gli elementi della collocazione" (Fontenelle, 1998:197, traduzione mia). Un altro problema di queste risorse lessicografiche è l'assenza di indicazioni sulla frequenza dei collocati e sul loro impiego in tipologie testuali o domini specifici. Come già notava Fontenelle (1992, 1998), la selezione dei collocati in questi tipi di dizionari spesso è arbitraria, basata sull'intuizione del lessicografo, e non su criteri quali la frequenza nell'uso; inoltre collocati frequenti e rari sono messi sullo stesso piano: per uno studente straniero potrebbe risultare assai complicato capire quali siano più comuni o adatti in determinati contesti.

### 4.2.3 Risorse per l'apprendimento delle collocazioni: i corpora

Per quanto riguarda quest'ultimo aspetto, ossia la possibilità di osservare i collocati di una parola e le rispettive frequenze, un aiuto significativo può venire dai corpora, la cui utilità nella didattica delle lingue e nella formazione di traduttori e interpreti è ormai riconosciuta (Bernardini e al. a cura di, 2003; Sinclair a cura di, 2004; Gallego-Hernandez e Rodriguez-Inés, 2016). Grazie all'introduzione dei corpora nei percorsi di formazione dedicati a traduttori e interpreti, gli studenti hanno a disposizione strumenti come i *concordancer* (cfr. cap. 3 sez. 3.2.5) per condurre autonomamente ricerche relative alle collocazioni di una data parola, ottenendo così collocati tratti da testi reali (e non frutto della selezione operata da un autore) corredati da dati sulle frequenze.

L'approccio basato sui corpora offre inoltre la possibilità di raffinare i criteri di ricerca: selezionando un corpus di lingua generale (cfr. sez. 3.2.4) è possibile ad esempio osservare le frequenze generali dei collocati in una data lingua, mentre restringendo il campo a corpora specialistici (cfr. sez. 3.2.4) è possibile osservare le frequenze d'uso di una data parola e quali collocati occorrono maggiormente in tipologie testuali o in domini tematici specifici.

Se i corpora possono costituire un valido aiuto allo studente per orientarsi tra le frequenze dei collocati, resta però aperto il problema legato all'assenza di distinzioni semantiche: un corpus, specialmente se non annotato semanticamente, non fornisce allo studente straniero alcuna informazione sul legame semantico e lessicale tra una parola e i suoi collocati, né sulle differenze di senso tra i vari collocati: lo studente è obbligato, anche in questo caso, a consultare un dizionario di lingua.

### 4.2.4 Conclusioni sulle risorse passate in rassegna

Alla luce di quanto esposto finora è dunque possibile formulare un quesito analogo a quello di Fontenelle (1992), il quale si chiedeva se il metodo più efficace per avere accesso a collocazioni corredate da informazioni linguisticamente rilevanti fosse estrarle dai corpora o da dizionari di collocazioni.<sup>39</sup> Fontenelle (1992) arriva alla conclusione che i due metodi sono

---

<sup>39</sup> È necessario precisare che la riflessione di Fontenelle (1992) è relativa all'estrazione automatica di informazioni sulle collocazioni al fine di creare database lessicali elettronici. Riteniamo però che le sue riflessioni siano applicabili in maniera estremamente pertinente anche all'ambito dell'apprendimento.

complementari, conclusione sostanzialmente condivisa da chi scrive, seppur con alcune riserve. Per avere informazioni complete la cosa migliore è effettivamente utilizzare più strumenti (dizionari di collocazioni, corpora e dizionari generali), ma ciò comporta la necessità di un cospicuo investimento di tempo da parte dello studente. Inoltre, vista l'assenza di una risorsa completa e di rapida consultazione, lo studente che desideri avere accesso a informazioni complete sulle collocazioni deve mettere a punto strategie autonome, come la costruzione di database personali.

Per questi motivi abbiamo avvertito la necessità di uno strumento nuovo che faciliti gli studenti nell'apprendimento delle collocazioni.

#### **4.2.5 La proposta oggetto del presente studio di caso**

A livello teorico riteniamo che le funzioni lessicali proposte nell'ambito della teoria Senso-Testo possano costituire un valido aiuto nell'apprendimento delle collocazioni, poiché come già esposto nel capitolo 2, permettono di rappresentare in modo sistematico la relazione semantica e lessicale esistente tra base e collocato, permettendo così di classificare i collocati di una data parola anche in base al loro senso.

Il metodo che abbiamo scelto per implementare le funzioni lessicali in uno strumento pratico è un'annotazione di corpus (cfr. cap. 3 sez. 3.3), che consiste nell'assegnare a ciascun collocato di una data parola un'etichetta con la funzione lessicale corrispondente. Il corpus così annotato andrebbe a costituire una risorsa per l'apprendimento delle collocazioni dotata delle seguenti caratteristiche:

- in quanto corpus, si tratta di una risorsa basata su dati testuali reali, su cui è possibile condurre analisi statistiche sulla frequenza dei collocati e sulla loro distribuzione;
- fornisce, grazie all'annotazione per funzioni lessicali, informazioni semantiche sui collocati e informazioni sulla natura del legame tra base e collocati;
- permette di concentrare tutte le informazioni rilevanti in un'unica risorsa;
- incoraggia lo studente ad analizzare autonomamente i dati raccolti e a riflettere sui risultati anziché accettare passivamente le informazioni fornite da strumenti come i dizionari.

Nelle sezioni successive descriveremo nel dettaglio il nostro studio di caso, a partire dal corpus sul quale è stato realizzato.

### 4.3 Il corpus EPTIC

Il corpus scelto per realizzare l'annotazione è EPTIC (*European Parliament Translation and Interpreting Corpus*), corpus trilingue realizzato a partire dall'anno 2013 presso il Dipartimento di Traduzione e Interpretazione dell'Università di Bologna sotto la direzione di Silvia Bernardini, Adriano Ferraresi e Maja Miličević Petrović. Si tratta di un'evoluzione del corpus EPIC (*European Parliament Interpreting Corpus*) creato tra il 2004 e il 2006 da studiosi dello stesso dipartimento e costituito da testi di interventi al Parlamento Europeo e delle relative interpretazioni in lingua inglese e spagnola (Sandrelli e Bendazzoli, 2005).

EPTIC presenta alcune importanti differenze rispetto a EPIC: oltre a sfruttare combinazioni linguistiche differenti (italiano, inglese e francese), EPTIC comprende anche le traduzioni scritte degli interventi.

EPIC e EPTIC, in quanto corpora di testi del parlamento europeo, rientrano in una famiglia di corpora che nell'ultimo decennio si è assai diffusa e ha trovato interessanti applicazioni in linguistica computazionale, in linguistica dei corpora e in studi sulla traduzione e l'interpretazione (Bernardini e al., 2016). I materiali dal Parlamento Europeo possiedono infatti caratteristiche molto interessanti ai fini della costruzione di corpora: sono testi autentici, tipologicamente omogenei, pubblicamente accessibili e disponibili in un grande numero di lingue (Bernardini e al., 2016).

La scelta del corpus EPTIC per questo studio di caso è motivata da due fattori principali:

- i testi istituzionali europei sono una tipologia testuale sempre più utilizzata nella formazione di traduttori e interpreti (Sosoni, 2011:99), poiché un'importante sbocco professionale di tali percorsi di formazione è precisamente la traduzione e l'interpretazione in contesti istituzionali;
- dal momento che il tipo di annotazione proposta presuppone una serie di operazioni da realizzare manualmente tra cui la disambiguazione semantica della base delle collocazioni e la cernita dei collocati (cfr. sez. 4.3.2), le dimensioni ridotte del corpus EPTIC consentono di realizzare un'annotazione completa, che riguardi cioè tutte le

occorrenze delle collocazioni prese in esame. L'annotazione manuale di corpora di grandi dimensioni richiederebbe un'enorme quantità di tempo e risorse, e non è affrontabile nel contesto di una tesi magistrale.

#### 4.3.1 Caratteristiche e struttura di EPTIC

La presente sezione è in larga parte basata sulle informazioni contenute in Ferraresi e Bernardini (in stampa). Come anticipato nella sezione 3.2.4, EPTIC è un corpus *specialistico*, perché contiene testi rappresentativi di un unico tipo testuale, ossia testi argomentativi di ambito politico e di carattere monologico (Stragapede, 2017). È un corpus *intermodale*, perché è costituito da testi originali e dalle rispettive traduzioni e interpretazioni. È un corpus *sincronico*: i testi di cui è composto sono tutti interventi della sessione plenaria della settima legislatura del Parlamento Europeo, tenutasi a Strasburgo tra il 17 e il 20 gennaio 2011. È un corpus trilingue contenente subcorpora in inglese, francese e italiano. Per quanto riguarda la direzionalità, EPTIC può essere considerato bidirezionale solo per le combinazioni linguistiche inglese<>italiano e inglese<>francese, poiché fino al 2011 la traduzione di interventi al Parlamento Europeo avveniva tramite la mediazione dell'inglese come lingua pivot (Stragapede, 2017). Ciò significa che ogni testo in lingua diversa dall'inglese veniva prima tradotto in inglese, e poi nelle altre lingue dell'Unione.

Per tali combinazioni linguistiche EPTIC funziona sia come corpus *parallelo*, poiché contiene testi originali con le relative traduzioni, che come corpus *comparabile*, poiché i testi originali in lingue diverse appartengono alla medesima tipologia testuale (cfr. cap. 3 sez. 3.2.4). Malgrado non si tratti di un corpus generale, rappresentativo della lingua nel suo complesso, e malgrado le dimensioni contenute (circa 330.000 parole), EPTIC è un corpus interessante per le sue applicazioni nello studio della traduzione e dell'interpretazione e in linguistica contrastiva.

Oltre a rappresentare una buona fonte di dati testuali, EPTIC è stato arricchito di metadati portatori di informazioni relative a ciascun testo (data e ora di produzione, durata, modalità di esecuzione, argomento ecc.), all'oratore (nome, nazionalità, genere) e all'interprete che ha prodotto la resa in lingua diversa dall'originale (parlante nativo o meno, genere ecc.). Il corpus è stato annotato a più livelli: oltre all'annotazione automatica per parti del discorso è stata

realizzata anche la lemmatizzazione mediante lo strumento TreeTagger.<sup>40</sup> Non sono tuttavia presenti altre annotazioni sintattiche o semantiche. Il corpus è codificato in formato CWB-compliant<sup>41</sup> ed è stato indicizzato in modo da consentirne l'accesso e l'esplorazione con NoSketch Engine (Cfr. cap. 3 sez. 3.2.5). Infine, EPTIC è stato ulteriormente arricchito mediante un paziente lavoro di allineamento text-to-text (ossia tra testi e traduzioni) e text-to-video (ossia l'allineamento tra testo scritto e file video dell'intervento), moltiplicando così le possibilità di esplorazione e sfruttamento del corpus.

La struttura attuale di EPTIC è composta da 14 sottocorpus, le cui caratteristiche sono riassunte nella tabella 1 (tratta da Ferraresi e Bernardini, in stampa).

Language	Sub-corpus	N. of texts	Word count
English	st-in-en	64	21,106
	st-tr-en	64	20,091
	tt-in-en_from-fr	65	20,836
	tt-tr-en_from-fr	65	21,641
	tt-in-en_from-it	68	16,122
	tt-tr-en_from-it	68	17,775
	<i>Sub-total</i>	394	117,571
French	st-in-fr	65	23,943
	st-tr-fr	65	23,159
	tt-in-fr_from-en	63	20,558
	tt-tr-fr_from-en	64	22,294
	<i>Sub-total</i>	257	89,954
Italian	st-in-it	68	16,586
	st-tr-it	68	16,521
	tt-in-it_from-en	64	16,977
	tt-tr-it_from-en	64	19,665
	<i>Sub-total</i>	264	69,749
TOTAL	1179	327,023	

Tabella 1: informazioni sulle dimensioni di EPTIC (Ferraresi e Bernardini, in stampa).

<sup>40</sup> Per informazioni su TreeTagger si veda <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>.

<sup>41</sup> Per informazioni su CWB (Corpus WorkBench) si veda <http://cwb.sourceforge.net/>.

Come si nota dalla tabella, ciascun sottocorpus è identificato mediante un nome che ne riassume le caratteristiche principali, strutturato nel modo seguente:

- st = *source text*, ossia un testo originale
- tt = *target text*, ossia un testo che è una traduzione o interpretazione di un testo originale
- tr = *traduzione*, etichetta usata per:
  - testi tradotti
  - testi originali (o meglio i *verbatim report* messi a disposizione dal Parlamento, ossia la trascrizione ufficiale di discorsi, “ripulita” da pause, ripetizioni, riformulazioni e altri marcatori dell’oralità) destinati a essere tradotti
- in = *interpretazione*, etichetta usata per:
  - testi interpretati
  - testi originali destinati a essere interpretati, trascritti includendo pause e altri marcatori dell’oralità, secondo convenzioni descritte in Santandrea (2014) e Ferraresi e Bernardini (in stampa)
- en = inglese
- fr = francese
- it = italiano

#### 4.3.2 Selezione del materiale linguistico da annotare

Come già affermato nella sezione 4.2, il presente studio di caso è una proposta metodologica riguardante una possibile applicazione delle funzioni lessicali nell’annotazione di corpora. Per questo motivo, anziché realizzare un’annotazione completa delle collocazioni del corpus EPTIC, che va oltre gli scopi dello studio di caso, abbiamo scelto di restringere il campo concentrandoci su una sola lingua e un solo lessema.

L’annotazione riguarda i quattro sottocorpora in lingua italiana, includendo sia testi originali che traduzioni e interpretazioni. La scelta della lingua italiana è motivata dal fatto che il tipo di annotazione che si intende realizzare presuppone un’eccellente competenza linguistica, perché per una corretta annotazione dei collocati è necessario saper operare alcune distinzioni semantiche molto sottili. L’italiano è la lingua in cui chi scrive sente di essere più competente per operare tali distinzioni.

I sottocorpora in lingua italiana non sono stati annotati per intero: il presente studio di caso è limitato alle collocazioni che riguardano il lessema DIRITTO. La possibile situazione d'uso dell'annotazione immaginata per questo studio di caso è la situazione in cui uno studente non madrelingua voglia usare il corpus EPTIC per osservare alcuni esempi d'uso dei collocati del sostantivo DIRITTO in testi istituzionali in lingua italiana.

Il motivo per cui si è scelto il lessema DIRITTO per lo studio di caso è che si tratta di un sostantivo relativamente frequente nei sottocorpora italiani di EPTIC e che presenta collocati interessanti ai fini dell'annotazione. La selezione del lessema DIRITTO è stata condotta in base ai criteri illustrati di seguito.

Innanzitutto sono stati definiti due criteri di selezione globale delle collocazioni annotabili, in modo che lo studio di caso potesse presentare un campione di funzioni lessicali interessanti:

- la base della collocazione deve essere un *lessema* e nello specifico un *sostantivo*, poiché molte funzioni lessicali interessanti corrispondono a collocazioni che hanno come base un sostantivo (Mel'čuk, 1998);
- per assicurarsi di lavorare su un campione sufficientemente ampio, la base della collocazione deve essere relativamente frequente in ciascun sottocorpus. Sinclair (2005) suggerisce che "A word which is not specially ambiguous will require at least twenty instances for even an outline description of its behaviour to be compiled by trained lexicographers", tuttavia in ragione delle dimensioni ridotte del corpus si è deciso di fissare due soglie arbitrarie di frequenza minima:  $\geq 10$  occorrenze in almeno uno dei sottocorpora e  $\geq 40$  occorrenze totali.

Si è quindi proceduto all'esplorazione dei sottocorpora e all'individuazione di sostantivi (cioè le potenziali basi delle collocazioni) tramite lo strumento *wordlist* del software *AntConc*, il quale fornisce una lista in ordine decrescente dei token più frequenti. I sostantivi più frequenti sono stati raccolti in una tabella, corredati delle rispettive occorrenze (calcolate sulle occorrenze del lessema, sommando cioè le occorrenze delle forme singolari e plurali). Alla luce di questa prima selezione sono emersi come possibili candidati i sostantivi riportati in Tabella 2:

Sostantivo	N. occorrenze totali nei sottocorpora di italiano
commissione	320
unione	310
presidente	302
accordo	269
diritto	179
parte	145
mercato	106
problema	97
obiettivo	67
impegno	61
esempio	56
libertà	51
discussione	50
norma	49
integrazione	42

Tabella 2: lessemi (sostantivi) risultanti dalla prima selezione del materiale linguistico.

Attraverso lo strumento *concordances* di *AntConc* sono state osservate le concordanze di ciascun sostantivo precedentemente selezionato ed è stata operata un'ulteriore selezione sulla base di criteri qualitativi come:

- **varietà di contesti:** ciascun sostantivo dovrebbe presentare una certa varietà di collocati, in modo da riflettere un'ampia varietà di funzioni lessicali. Per questo motivo sono stati scartati i sostantivi che occorrono in una piccola varietà di contesti ripetitivi, come ad esempio *unione* (di cui si contano 259 co-occorrenze con *europa* su 320 occorrenze totali, pari all'80,93%) o *presidente* (nel 39% dei casi associato al collocato *Signor*) e che pertanto non sono interessanti per i nostri scopi;
- **presenza di collocati verbali:** i collocati verbali spesso corrispondono a funzioni lessicali interessanti, ad esempio i *verbi di supporto* o i cosiddetti *verbi di realizzazione* (cfr. cap. 2 sez. 2.5.1 e sez. 2.5.2). Per questo motivo sono stati selezionati in via preferenziale i sostantivi che presentano almeno un 50% di collocati verbali sul totale dei collocati;
- **presenza di collocazioni nel senso inteso da Mel'čuk:** sono stati scartati i sostantivi che presentano un elevato numero di co-occorrenze che non sono *collocazioni* nel

senso definito da Mel'čuk ma *libere associazioni tra parole* (per questa distinzione si veda cap. 2 sez. 2.3.2);

- **corrispondenza con FL standard:** le collocazioni, per poter essere annotate, devono essere rappresentabili in termini di funzioni lessicali standard (cfr. cap. 2 sez. 2.4.2.1); per questo motivo sono stati scartati i sostantivi che presentano una maggioranza di collocazioni che non corrispondono a funzioni lessicali standard (ad esempio *accordo bilaterale* è una collocazione, ma non corrisponde ad alcuna FL standard).

Alla luce di questa ulteriore selezione abbiamo infine optato per il sostantivo DIRITTO, che presenta tutte le caratteristiche appena elencate (v. Tabella 3 relativamente alla frequenza del sostantivo nei quattro sottocorpora). Occorre specificare che, data l'assenza di annotazione semantica, è stato necessario realizzare una disambiguazione manuale tra le due principali accezioni di DIRITTO e di conseguenza un affinamento del calcolo delle occorrenze, che ha dato come esito 26 occorrenze dell'accezione "insieme di norme" e 153 occorrenze del lessema nell'accezione "facoltà, prerogativa". Si è quindi deciso di incentrare lo studio di caso sul lessema DIRITTO nella sua seconda accezione, indicata da ora in poi come DIRITTO<sub>2</sub>, che possiede 73 occorrenze di collocati corrispondenti a funzioni lessicali standard.

sottocorpus	st-tr-it	st-in-it	tt-tr-it_from-en	tt-in-it_from-en
n. occorrenze	38	38	42	35

Tabella 3: frequenze di DIRITTO<sub>2</sub> nei sottocorpora di EPTIC

La prossima sezione illustra come i collocati così individuati sono stati annotati.

#### 4.4 Annotazione per funzioni lessicali

Come già accennato, l'annotazione oggetto di questo studio di caso riguarda le collocazioni del lessema DIRITTO<sub>2</sub> nei sottocorpora in lingua italiana. Dal punto di vista pratico ciò consiste nell'assegnazione, a ciascun collocato del lessema diritto, di un'etichetta con la funzione lessicale a cui ciascun collocato corrisponde.

Nelle sezioni a seguire saranno illustrate le caratteristiche e le modalità di realizzazione dell'annotazione.

#### 4.4.1 Schema di annotazione

In questa sezione introdurremo lo schema di annotazione (Leech, 2005) che, come illustrato nel capitolo 3 (sezione 3.3.1), serve a caratterizzare l'annotazione da un punto di vista strettamente linguistico, specificando:

- di che tipo di annotazione si tratta
- a quale teoria linguistica si fa riferimento
- la lista delle etichette o tag inseriti (detta anche *tagset*, termine che utilizzeremo da qui in poi) e a quali categorie linguistiche corrispondono
- *le annotation practices* (Leech, 2005), ossia i criteri con cui i tag sono stati associati ai collocati nel corpus.

Come già accennato nel capitolo 3 (sezione 3.3.3.5) l'annotazione proposta non è assimilabile alle annotazioni linguistiche "classiche", poiché non corrisponde a un singolo livello di analisi linguistica, ma piuttosto a un singolo fenomeno, le collocazioni. L'annotazione proposta permette di codificare al contempo sia informazioni sintattiche relative a ciascun collocato (ad esempio, in caso di collocati verbali, alcune informazioni sugli attanti) sia, soprattutto, informazioni semantiche, permettendo così di differenziare i collocati in base al loro senso.

La teoria linguistica a cui si fa riferimento è la teoria Senso-Testo (cfr. cap.1), in particolare le funzioni lessicali teorizzate da Mel'cuk (cfr. cap. 2, sez. 2.4).

Trattandosi di un'annotazione originale ideata da chi scrive, il tagset non fa riferimento ad alcuno standard, ma è stato ideato e costruito tenendo presente come unici criteri quelli esposti in Leech (2005), secondo cui i tag devono possedere tre caratteristiche principali: *brevità*, *univocità* (ciascun tag deve cioè rappresentare una e una sola categoria in modo non ambiguo) e *trasparenza* (i tag devono essere cioè intelligibili per l'utente umano, cioè formati da sigle o abbreviazioni, e non da "stringhe arbitrarie di caratteri come Q!@" (Leech, 2005:21). In un primo momento si è pensato di costruire tag ispirati alla *codifica formale esplicita* proposta da Kahane e Polguère (2001), che costituisce uno standard per la codifica delle funzioni lessicali utilizzato in varie applicazioni computazionali (Jousse, 2010; Lambrey, 2016). Questa codifica, strutturata come una matrice, permette di rappresentare esplicitamente informazioni riguardanti la struttura semantica e sintattica della FL, in

particolare gli attanti sintattici. Ecco come si presenterebbe la codifica esplicita della funzione lessicale Oper<sub>1</sub>(diritto):

$$\left( \begin{array}{c} \textit{diritto} \\ V[1, \textit{diritto}] \end{array} \right)$$

ed ecco come si presenterebbe il tag per la funzione lessicale Oper<sub>1</sub> ispirato a tale codifica esplicita:

$$([\textit{diritto}], v[1, \textit{diritto}])$$

I tag così costruiti risultano eccessivamente lunghi e poco intelligibili per l'utente umano (un grande problema è rappresentato dal fatto che nella codifica esplicita non compare il nome della funzione lessicale). Quest'idea è stata dunque scartata, e si è preferito optare per tag strutturati diversamente.

I tag ideati per l'annotazione coincidono con i nomi delle varie funzioni lessicali. Tale scelta è motivata dal fatto che i nomi delle funzioni lessicali possiedono già caratteristiche di univocità e brevità. Gli indici numerici presenti in alcune funzioni lessicali sono stati inclusi, senza tuttavia la possibilità di differenziare indici in apice o in pedice:

**Esempio di grafia originale**

Real<sub>4</sub><sup>II</sup>

**Esempio di tag (grafia senza indici in apice/pedice)**

Real4II

#### 4.4.1.1 Tagset dell'annotazione

Il tagset ideato per l'annotazione è qui presentato, conformemente a quanto suggerito da Leech (2005), in un glossario tabulare in cui per ciascun tag viene fornita la funzione lessicale corrispondente, una breve spiegazione e un esempio tratto dal corpus EPTIC riguardante un collocato di DIRITTO<sub>2</sub>. Come affermato da Leech (2005:22), il tagset in forma di glossario funziona idealmente come un "convenient first port of call for people trying to make sense of the annotation". Nei paragrafi successivi le spiegazioni di ciascuna funzione lessicale sono riprese solo in sintesi. Per spiegazioni più approfondite sulle funzioni lessicali si rimanda al capitolo 2 e alla tabella delle funzioni lessicali in Appendice 2.

▪ Oper<sub>1</sub>

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
Oper <sub>1</sub>	Oper <sub>1</sub>	<p>Corrisponde a un verbo di supporto (ossia “vuoto” semanticamente) il cui soggetto è il primo attante semantico della base (cioè, nel caso del lessema DIRITTO, il titolare del diritto).</p> <p>La base della collocazione (DIRITTO) costituisce invece il secondo argomento obbligatorio del verbo.</p>	<b>Avere</b> (“Ogni Paese ha diritto a scegliersi la propria strada”).

▪ Real<sub>1</sub>

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
Real <sub>1</sub>	Real <sub>1</sub>	<p>Corrisponde a un <i>verbo di realizzazione</i> (Mel'čuk, 1996), che esprime cioè un senso di <i>realizzazione</i> associato a DIRITTO. Beneficiare di un diritto significa “farlo realizzare”, realizzare ciò per cui un diritto è stato concepito: ecco in che cosa consiste il senso di realizzazione.</p> <p>Nell'esempio a fianco “i pazienti beneficiano di diritti già riconosciuti” corrisponde a “i pazienti fanno ciò che devono fare con i propri diritti”.</p> <p>Il soggetto del verbo è il primo attante semantico della base (cioè, nel caso del lessema DIRITTO, il titolare del diritto), espresso dall'indice 1 in cifra araba.</p> <p>Anche nel caso della funzione Real la base della collocazione (cioè DIRITTO) costituisce il secondo argomento obbligatorio del verbo.</p>	<b>Beneficiare</b> (“questo permetterà ai pazienti di beneficiare di diritti che sono già riconosciuti”).

▪ Real<sub>3</sub>

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
Real <sub>3</sub>	Real <sub>3</sub>	<p>Corrisponde a un verbo dotato di un senso di <i>realizzazione</i> (vedi Real<sub>1</sub>). Si tratta di un verbo di realizzazione perché esprime la realizzazione del ruolo dell'attante rispetto al diritto in questione: nell'esempio a fianco gli</p>	<b>Garantire</b> (“gli Stati membri devono garantire il diritto all'adozione”).

		<p>Stati “fanno ciò che è loro compito fare” rispetto al diritto citato.</p> <p>Il soggetto del verbo è in questo caso il <i>terzo</i> attante semantico della base (per il lessema DIRITTO, l'autorità che stabilisce un diritto), espresso dall'indice 3 in cifra araba.</p>	
--	--	--	--

▪ **Real<sub>41</sub>**

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
Real <sub>41</sub>	Real <sub>41</sub>	<p>Corrisponde a un verbo che esprime un senso di <i>realizzazione</i>. Il verbo esprime la realizzazione del ruolo del <i>quarto</i> attante della base, cioè gli individui di fronte a cui un diritto viene fatto valere, e che sono tenuti a riconoscerlo. “riconoscere un diritto a qualcuno” corrisponde a “fare ciò che si è tenuti a fare” con un diritto altrui (cfr. esempio a fianco).</p> <p>Il soggetto del verbo è il <i>quarto</i> attante semantico della base, la base (DIRITTO) è il secondo argomento obbligatorio <u>ed è inoltre presente un terzo argomento obbligatorio</u>, rappresentato dal primo attante semantico di DIRITTO (cioè il titolare del diritto).</p> <p>Il motivo per cui sono presenti due numeri arabi in pedice è appunto che essi corrispondono agli attanti della base che devono necessariamente essere espressi (in quanto argomenti obbligatori).</p> <p>Per approfondimenti cfr. cap. 2 sez. 2.5.2.1, Appendice 2 e Appendice 3.</p>	<p><b>Riconoscere a X un diritto</b> (“si è cercato di riconoscere a tutti i pazienti il diritto all'assistenza sanitaria”)</p>

▪ **AntiReal<sub>41</sub>**

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
AntiReal <sub>41</sub>	AntiReal <sub>41</sub>	<p>Funzione lessicale complessa. Corrisponde a un verbo il cui senso è “fare il contrario di ciò che si è tenuti a fare con un diritto altrui”.</p> <p>Il verbo in questione indica cioè il contrario del verbo espresso da Real<sub>41</sub>, il contrario di ciò che il <i>quarto</i> attante semantico della base dovrebbe fare con un diritto altrui. L'assetto sintattico è lo stesso di Real<sub>41</sub>.</p>	<p><b>Negare</b> (“questo diritto è stato negato ai cristiani”)</p>

▪ Real<sub>4</sub><sup>II</sup>

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
Real <sub>4</sub> <sup>III</sup>	Real <sub>4</sub> <sup>II</sup>	<p>Corrisponde a un verbo che esprime un senso di <i>realizzazione</i>. Il verbo esprime infatti la realizzazione del ruolo dell'attante rispetto al diritto in questione: nell'esempio a fianco la democrazia "fa ciò che è tenuta a fare" con i diritti fondamentali.</p> <p>In questo caso l'indice II in cifre romane in apice serve a segnalare una diversa sfumatura di senso, un maggiore "grado di realizzazione" rispetto a Real<sub>4</sub><sup>I</sup> (<i>riconoscere</i>, assente nel presente tagset, per cui si rimanda alla spiegazione contenuta in Appendice 3). "Rispettare un diritto" corrisponde infatti a un maggior grado di realizzazione rispetto a "riconoscere un diritto". Si veda anche la spiegazione contenuta nella sezione 2.5.2.1.</p> <p>Il soggetto del verbo è il <i>quarto</i> attante semantico della base (nel caso del lessema DIRITTO, coloro che sono tenuti al rispetto di un diritto), mentre la base (DIRITTO) è il secondo argomento obbligatorio.</p>	<p><b>Rispettare</b> ("la Tunisia vuole una democrazia stabile che rispetti pienamente i diritti [...] fondamentali")</p>

▪ AntiReal<sub>4</sub><sup>II</sup>

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
AntiReal <sub>4</sub> <sup>II</sup>	AntiReal <sub>4</sub> <sup>II</sup>	<p>Funzione lessicale complessa. Corrisponde a un verbo che esprime il contrario del verbo espresso da Real<sub>4</sub><sup>II</sup>: se Real<sub>4</sub><sup>II</sup> corrisponde a "rispettare", corrisponderà al suo contrario, cioè "non rispettare", "violare", in altre parole "fare il contrario di ciò che si dovrebbe fare con un diritto altrui".</p> <p>L'indice II in cifra romana indica a sfumatura di senso diversa da AntiReal<sub>4</sub><sup>I</sup> ("non riconoscere", non presente nel tagset): "non rispettare", "violare un diritto" ha una sfumatura più forte rispetto a "non riconoscere" (cfr. cap. 2 sez. 2.5.2.1). L'assetto sintattico è lo stesso di Real<sub>4</sub><sup>II</sup>.</p>	<p><b>Violare</b> ("questo diritto è stato violato nel mio paese")</p>

▪ **AntiReal<sub>4</sub><sup>III</sup>**

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
AntiReal <sub>4</sub> <sup>III</sup>	AntiReal <sub>4</sub> <sup>III</sup>	<p>Questa funzione lessicale complessa corrisponde a un verbo che esprime “il contrario di ciò che si dovrebbe fare con un diritto altrui”, ma con una sfumatura di senso diversa sia da AntiReal<sub>4</sub><sup>I</sup> che da AntiReal<sub>4</sub><sup>II</sup>: il senso di “intaccare i diritti” o “ledere i diritti” implica infatti un comportamento lesivo a lungo termine, e non un danno “puntuale” come nel caso di una violazione, che può costituire un singolo episodio.</p> <p>La diversità di sfumatura è segnalata dall’indice III in numeri romani (cfr. cap. 2 sez. 2.5.2.1).</p>	<b>Intaccare</b> (“un contratto collettivo che intacca i diritti fondamentali dei lavoratori”)

▪ **S<sub>0</sub>Real<sub>4</sub><sup>II</sup>**

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
S <sub>0</sub> Real <sub>4</sub> <sup>II</sup>	S <sub>0</sub> Real <sub>4</sub> <sup>II</sup>	<p>Questa funzione lessicale complessa corrisponde a un sostantivo dal significato “ciò che bisognerebbe fare con un diritto altrui”.</p> <p>Il senso del collocato nominale è dunque un senso di realizzazione.</p> <p>Il collocato costituisce una sostantivazione (indicata dalla FL S<sub>0</sub>) del verbo espresso da Real<sub>4</sub><sup>II</sup>.</p>	<b>Rispetto</b> (“l’accordo contempla il rispetto dei diritti [...] delle popolazioni indigene”)

▪ **Adv<sub>0</sub>Real<sub>4</sub><sup>II</sup>**

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
Adv <sub>0</sub> Real <sub>4</sub> <sup>II</sup>	Adv <sub>0</sub> Real <sub>4</sub> <sup>II</sup>	<p>Questa funzione lessicale complessa corrisponde a un’espressione avverbiale che significa “in un modo tale da rispettare un diritto altrui”.</p> <p>Si nota che il senso di realizzazione è analogo a quello di Real<sub>4</sub><sup>II</sup>: l’espressione è in effetti ottenuta per derivazione sintattica (indicata dalla FL Adv<sub>0</sub>) dal verbo espresso da Real<sub>4</sub><sup>II</sup>.</p>	<b>Nel rispetto di</b> (“garantire che l’adozione si svolga [...] nel rispetto dei diritti”)

▪ S<sub>0</sub>AntiReal<sub>4</sub><sup>II</sup>

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
S <sub>0</sub> AntiReal <sub>4</sub> <sup>III</sup>	S <sub>0</sub> AntiReal <sub>4</sub> <sup>II</sup>	<p>Questa funzione lessicale complessa corrisponde a un sostantivo il cui senso è “ciò che NON bisognerebbe fare con un diritto altrui”.</p> <p>Il collocato costituisce una sostantivazione (indicata da S<sub>0</sub>) del verbo espresso da AntiReal<sub>4</sub><sup>II</sup>.</p>	<b>Violazione</b> (“è un caso di violazione dei diritti dell’uomo”)

▪ PermFunc<sub>0</sub>

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
PermFunc <sub>0</sub>	PermFunc <sub>0</sub>	<p>Questa funzione lessicale complessa corrisponde a un verbo il cui senso è “causare la continuazione dell’esistenza di un diritto”.</p> <p>Il soggetto di tale verbo può essere un agente esterno rispetto agli attanti semantici del lessema in questione.</p>	<b>Tutelare</b> (“il ruolo di chi dovrebbe tutelare i diritti dei minori...”)

▪ S<sub>0</sub>PermFunc<sub>0</sub>

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
S <sub>0</sub> PermFunc <sub>0</sub>	S <sub>0</sub> PermFunc <sub>0</sub>	<p>Questa funzione lessicale complessa corrisponde a un sostantivo il cui senso è “il fatto di causare la continuazione dell’esistenza di un diritto”.</p> <p>Il collocato costituisce una sostantivazione (indicata da S<sub>0</sub>) del verbo espresso da PermFunc<sub>0</sub>.</p>	<b>Tutela</b> (“è fondamentale prestare maggiore attenzione alla tutela dei diritti”)

▪ S<sub>1</sub>PermFunc<sub>0</sub>

Tag	Funzione lessicale	Spiegazione	Esempio di collocato e di uso
S <sub>1</sub> PermFunc <sub>0</sub>	S <sub>1</sub> PermFunc <sub>0</sub>	<p>Questa funzione lessicale complessa corrisponde a un sostantivo il cui senso è “l’agente che causa la continuazione dell’esistenza di un diritto”.</p>	<b>Difensore</b> (“l’Europa, storicamente [...] grande difensore dei diritti umani”)

		Il collocato non costituisce una semplice sostantivazione del verbo espresso da PermFunc0: si tratta più precisamente di un sostantivo che rappresenta l'agente del verbo (cfr. funzione lessicale S <sub>1</sub> in Appendice 2).	
--	--	--	--

Per quanto riguarda i criteri con cui i tag sono stati associati ai collocati di DIRITTO<sub>2</sub>, occorre una precisazione. Una categorizzazione rigorosa dei collocati dovrebbe fare riferimento a uno strumento come un Dizionario esplicativo-combinatorio (cfr. cap.1 sez. 1.6) dell'italiano, che indichi tutte le funzioni lessicali pertinenti per il lessema DIRITTO<sub>2</sub> e i valori di tali funzioni lessicali. Purtroppo, come già accennato, tale dizionario per la lingua italiana non esiste.

Per ovviare a questo problema è stata realizzata una scheda lessicografica del lessema italiano DIRITTO<sub>2</sub> (cfr. Appendice 3) compilata in base ai criteri del DEC francese, descritti in Mel'čuk e al. (1995; 1984-1999) e utilizzando come parziale riferimento il DiCoLiLex, un database elettronico di collocazioni del francese attualmente in corso di realizzazione all'Université de Montréal (Marengo e Robichaud, approvato). Tali strumenti sono stati solo parzialmente utili, perché nonostante il francese e l'italiano siano lingue tipologicamente simili, esistono importanti differenze che rendono impossibile applicare all'italiano la categorizzazione utilizzata per il francese. Per la realizzazione della scheda è stato dunque necessario un paziente lavoro di analisi semantica e sintattica del lessema DIRITTO<sub>2</sub> e delle sue collocazioni (estratte dal corpus EPTIC e in parte dai dizionari di Urzì e Tiberii), svolto da chi scrive in stretta collaborazione con il professor Lareau dell'Université de Montréal.

La scheda contiene informazioni esplicite sulla struttura di attanti semantici del lessema DIRITTO<sub>2</sub> e un elenco di funzioni lessicali, con il collocato (o i collocati) corrispondente al valore di ciascuna funzione lessicale (sul modello del DEC). La scheda così elaborata ha costituito il riferimento per poter successivamente annotare il corpus associando a ogni collocato il tag con la funzione lessicale pertinente, e per la costruzione del tagset presentato nella sezione 4.4.1.1.

#### 4.4.2 Schema di codifica

In questa sezione introdurremo lo schema di codifica (Leech, 2005) che, come illustrato nel capitolo 3 (sezione 3.3.1), serve a illustrare come l'informazione linguistica è stata codificata, ossia il modo in cui l'annotazione è stata realizzata concretamente.

L'annotazione oggetto del presente lavoro è stata realizzata su sottocorpora di EPTIC precedentemente annotati tramite lo strumento TreeTagger (Cfr. sez. 4.3.1), il cui formato di codifica è un misto di XML e formato verticale CWB-compliant (Ferraresi e Bernardini, in stampa). Il formato verticale prodotto da TreeTagger (visibile in Figura 25) ha una struttura in cui ciascun token del corpus è riportato su una riga, e i tag relativi ai vari livelli di annotazione sono inseriti sulla stessa riga di ciascun token, in colonne diverse distanziate tramite *tab*.

L'annotazione è dunque stata codificata in formato verticale, mediante l'aggiunta di un'ulteriore colonna (denominata *lexf* per *Lexical Function*) alla struttura verticale (cfr. Figura 25).

Capisco	VER:fin	capire	
le	ART	la	
motivazioni	NOUN	motivazione	
,	PUN	,	
però	ADV	però	
vorrei	VER2:fin	volere	
dimostrare	VER:infi	dimostrare	
che	CHE	che	
non	NEG	non	
è	VER:fin	essere	
un	ART	un	
caso	NOUN	caso	
di	PRE	di	
violazione	NOUN	violazione	S0AntiReal4II_diritto_24
dei	ARTPRE	del	
diritti	NOUN	diritto	base_24
dell'	ARTPRE	dell'	
uomo	NOUN	uomo	

Figura 25: formato verticale del corpus EPTIC annotato.

L'annotazione è stata realizzata manualmente tramite l'editor di testo Notepad++. Non è stato possibile automatizzare l'annotazione a causa di diversi fattori, tra cui l'assenza di annotazione semantica e la conseguente necessità di disambiguare manualmente le occorrenze del lessema DIRITTO<sub>2</sub> per poter annotare solo quelle corrispondenti all'accezione "facoltà, prerogativa".

L'annotazione è stata realizzata sia sui collocati che sulla base della collocazione: i tag presentati nella sezione 4.4.1.1 sono stati utilizzati per i collocati di DIRITTO<sub>2</sub>, ma anche ciascuna occorrenza della base è stata annotata mediante il tag *base* seguito da un trattino basso e da un numero progressivo. Il numero progressivo serve a identificare in modo univoco i due elementi della collocazione: il tag di ciascun collocato reca infatti un numero progressivo corrispondente al numero del tag della base (cfr. Figura 25).

Chiudiamo la sezione con una precisazione. L'annotazione realizzata nel contesto di questo studio di caso è un'annotazione a tag singolo che permette di identificare le collocazioni del lessema DIRITTO<sub>2</sub> nei sottocorpora italiani di EPTIC; l'annotazione permette di identificare i due token che rappresentano gli elementi principali della collocazione (cioè il token che costituisce la base della collocazione e il token che ne costituisce il *collocato lessicale principale*). Naturalmente siamo consapevoli del fatto che le collocazioni spesso sono formate da più di due token: nella maggior parte dei casi individuati nei sottocorpora italiani di EPTIC, le collocazioni di *diritto* includono articoli, preposizioni o altri determinanti (come nel caso di *violazione dei diritti* in Figura 25) oppure verbi ausiliari. La scelta di annotare solo il collocato lessicale principale escludendo tutti i collocati che costituiscono parole grammaticali è stata dettata da un criterio di praticità nella consultazione. Qualora tutte le parole fossero state annotate, una ricerca tramite l'interfaccia di NoSketch Engine del tipo

input: funzione lessicale → output: collocato nel suo contesto di frase

avrebbe dato un risultato poco soddisfacente: i collocati non sarebbero comparsi su una riga sola, ma su tante righe quanti sono i collocati annotati:

Esempio di interrogazione (CQL)	lexf="S0AntiReal4II_**"
Output ideale desiderato (KWIC)	non è un caso di <b>violazione dei</b> diritti dell'uomo
Output dell'interrogazione se fossero stati annotati tutti i collocati	non è un caso di <b>violazione</b> dei diritti dell'uomo non è un caso di violazione <b>dei</b> diritti dell'uomo
Output ottenuto grazie all'attuale annotazione	non è un caso di <b>violazione</b> dei diritti dell'uomo

Per questo motivo si è deciso di annotare unicamente il collocato lessicale principale, in modo da non creare confusione nella visualizzazione dei risultati.

### 4.4.3 Modalità di consultazione dell'annotazione

L'annotazione così costruita, una volta indicizzata, è adatta a essere esplorata con NoSketch Engine. L'indicizzazione non è stata realizzata nel contesto di questo studio, dunque gli esempi di consultazione che andiamo a esporre hanno uno scopo puramente illustrativo.

Il corpus annotato può essere consultato in vari modi. Qualora l'utente desiderasse esplorare le concordanze della parola DIRITTO<sub>2</sub>, all'interno della finestra concordanze i collocati di DIRITTO<sub>2</sub> apparirebbero con il tag recante la funzione lessicale a cui essi corrispondono, permettendo così all'utente di differenziare i collocati per senso. I tag con le funzioni lessicali, qualora l'annotazione fosse indicizzata, potrebbero essere resi visibili mediante un'impostazione di NoSketch Engine analoga a quella che rende visibili i pos tag nella figura sottostante:

Query **diritto** 21 (1,052.26 per million)

Page 1 of 2 Go [Next](#) | [Last](#)

#305	ogni /ADJ paese /NOM ha /VER:pres il /DET:def <b>diritto</b> /NOM a /PRE scegliersi /VER:infi la /DET:def propria
#2971	Dobbiamo /VER:pres tutelare /VER:infi il /DET:def <b>diritto</b> /NOM del /PRE:det consumatore /NOM a /PRE conoscere
#3293	dobbiamo /VER:pres garantire /VER:infi il /DET:def <b>diritto</b> /NOM ad /PRE essere /VER:infi adottati /VER:ppper
#3299	/VER:infi adottati /VER:ppper , /PON il /DET:def <b>diritto</b> /NOM ad /PRE avere /VER:infi una /DET:indef famiglia
#3359	/DET:def bambini /NOM hanno /VER:pres il /DET:def <b>diritto</b> /NOM di /PRE essere /VER:infi adottati /VER:ppper
#3445	nel /PRE:det garantire /VER:infi il /DET:def <b>diritto</b> /NOM all' /PRE:det adozione /NOM e /CON per /PRE
#5073	/VER:geru al /PRE:det massimo /NOM il /DET:def <b>diritto</b> /NOM di /PRE adottare /VER:infi i /DET:def minore
#6318	/DET:def pieno /ADJ rispetto /NOM del /PRE:det <b>diritto</b> /NOM umanitario /ADJ internazionale /ADJ - /SENT
#6498	delle /PRE:det violazioni /NOM del /PRE:det <b>diritto</b> /NOM umanitario /ADJ internazionale /ADJ e /CON
#6679	/VER:pres ogni /ADJ giorno /NOM un /DET:indef <b>diritto</b> /NOM fondamentale /ADJ dell' /PRE:det umanità /NOM
#6906	/DET:def sicurezza /NOM alimentare /ADJ come /PRE <b>diritto</b> /NOM fondamentale /ADJ dell' /PRE:det umanità /NOM
#7064	/VER:pres definita /VER:ppper un /DET:indef <b>diritto</b> /NOM fondamentale /ADJ e /CON , /PON quindi /ADV
#7683	tutti /ADJ i /DET:def pazienti /NOM il /DET:def <b>diritto</b> /NOM all' /PRE:det assistenza /NOM sanitaria /ADJ
#8201	/VER:pres di /PRE affermare /VER:infi un /DET:indef <b>diritto</b> /NOM che /PRO:rela è /VER:pres quello /PRO:demo
#8396	un /DET:indef vero /ADJ e /CON proprio /ADJ <b>diritto</b> /NOM di /PRE scelta /NOM più /ADV consapevole /ADJ

Figura 26: visualizzazione KWIC di base e collocati in NoSketch Engine

Affinché l'annotazione risulti utile, l'utente deve possedere una conoscenza di base delle funzioni lessicali che gli consenta di capire a quale senso corrisponde ciascuna funzione lessicale. In questo senso il tagset dell'annotazione riportato nella sezione 4.4.1.1 può costituire un valido aiuto. Un'ulteriore strumento che può aiutare l'utente a orientarsi è la scheda del lessema DIRITTO<sub>2</sub> riportata in appendice al presente lavoro (cfr. Appendice 3).

Oltre all'esplorazione delle concordanze del lessema DIRITTO<sub>2</sub> sono possibili vari tipi di interrogazioni (*queries*), formulate secondo le regole della sintassi CQL oppure espresse in forma di espressioni regolari.

È possibile ad esempio visualizzare tutte le occorrenze di DIRITTO<sub>2</sub> che fanno parte di collocazioni corrispondenti a funzioni lessicali mediante la query `[lexf="base_.*"]`, oppure cercare tutti i collocati nominali di DIRITTO<sub>2</sub> mediante la query `[lexf="S.*_diritto_.*"]`.

È possibile inoltre fare ricerche semantiche mirate, ad esempio, a cercare tutti i collocati verbali di DIRITTO<sub>2</sub> che esprimono un senso di realizzazione, mediante la query `[lexf="Real.*_diritto_.*"]`.

#### 4.5 Conclusioni sullo studio di caso

L'annotazione proposta per questo studio di caso permette di arricchire un corpus integrando vari tipi di informazioni relative al fenomeno delle collocazioni, e in particolare offrendo la possibilità di differenziare per senso i collocati di un dato lessema.

I collocati annotati sono raccolti nella tabella 4 sottostante, raggruppati per funzione lessicale, assieme al numero di occorrenze di ciascuno nei vari sottocorpora.

Collocato	FL	n. occorrenze in st-in-it	n. occorrenze in st-tr-it	n. occorrenze in tt-in-it-from EN	n. occorrenze in tt-tr-it-from EN	TOT
Avere	Oper <sub>1</sub>	3	3	-	-	6
Beneficiare	Real <sub>1</sub>	2	2	-	-	4
Garantire	Real <sub>3</sub>	2	2	-	1	5
Riconoscere a X un diritto	Real <sub>41</sub>	1	1	-	1	3
Negare a X un diritto	AntiReal <sub>41</sub>	1	1	-	1	3
Riconoscere	Real <sub>4<sup>I</sup></sub>	2	2	-	-	4
Rispettare	Real <sub>4<sup>II</sup></sub>	1	1	3	3	8
Violare	AntiReal <sub>4<sup>II</sup></sub>	-	-	1	-	1
Intaccare	AntiReal <sub>4<sup>III</sup></sub>	1	1	-	-	2

Rispetto	SoReal <sub>4</sub> <sup>II</sup>	1	1	1	-	3
Nel rispetto	Adv <sub>0</sub> Real <sub>4</sub> <sup>II</sup>	1	1	-	1	3
Violazione	SoAntiReal <sub>4</sub> <sup>II</sup>	-	-	1	1	2
Difendere	PermFunc <sub>0</sub>	2	2	-	-	4
Proteggere	PermFunc <sub>0</sub>	-	-	1	1	2
Tutelare	PermFunc <sub>0</sub>	4	4	2	2	12
Difesa	SoPermFunc <sub>0</sub>	-	-	1	1	2
Tutela	SoPermFunc <sub>0</sub>	1	1	2	2	6
Difensore	S <sub>1</sub> PermFunc <sub>0</sub>	1	1	1	-	3

Tabella 4: collocati di DIRITTO<sub>2</sub> estratti dai sottocorpora italiani di EPTIC.

Nel processo di annotazione ci si è più volte confrontati con casi dubbi, in cui la corrispondenza tra collocato e funzione lessicale ha richiesto una riflessione più approfondita. Esistono ad esempio collocati verbali di DIRITTO<sub>2</sub> che, in contesti diversi, sembrano corrispondere a funzioni lessicali diverse. Uno di questi collocati verbali è *beneficiare*. *Beneficiare di un diritto* può voler dire, a seconda dei contesti, ‘essere titolare di un diritto’, cioè ‘avere un diritto’ (corrispondente alla funzione lessicale Oper<sub>1</sub>) oppure ‘esercitare un diritto’, cioè ‘fare ciò che un soggetto deve fare con un suo diritto’, in cui si ravvisa un senso di realizzazione che è rappresentabile mediante la funzione lessicale Real<sub>1</sub> (cfr. cap. 2 sez. 2.5.2.1). Nei sottocorpora italiani di EPTIC abbiamo trovato due sole occorrenze di questo collocato, che sembrano corrispondere più a questo secondo senso:<sup>42</sup>

“questo permetterà a tutti i pazienti [...] di **beneficiare** di diritti che sono già riconosciuti [...] nelle sentenze della Corte di giustizia dell’unione Europea”

“la direttiva da noi oggi approvata permetterà finalmente a tutti i pazienti di **beneficiare**, sull’intero territorio europeo, di un certo numero di diritti e prestazioni sanitarie”

Sia nel primo che nel secondo caso *beneficiare* sembra essere inteso come ‘usufruire di diritti’, cioè appunto un senso di realizzazione. Per questo motivo entrambe le occorrenze di *beneficiare* associate al lessema DIRITTO<sub>2</sub> sono state etichettate con Real<sub>1</sub>.

<sup>42</sup> Le occorrenze si trovano nei sottocorpora st-tr-it e st-in-it, che contengono gli stessi testi trascritti semplicemente con convenzioni diverse.

Un altro caso ambiguo è dato dal collocato verbale *tutelare*. In base ai criteri della lessicologia Senso-Testo (Mel'čuk e al., 1995), il senso di *tutelare* in rapporto al lessema DIRITTO<sub>2</sub> è analizzabile in due modi principali:

- (1) *tutelare un diritto*, specialmente se il soggetto del verbo *tutelare* coincide con il terzo attante semantico di DIRITTO<sub>2</sub> (ossia una legge o l'autorità garante di un diritto) può significare 'fare ciò che si deve fare con un diritto', ossia un senso di realizzazione del ruolo del terzo attante semantico rispetto a DIRITTO<sub>2</sub>. Questo senso corrisponde alla funzione lessicale Real<sub>3</sub> (cfr. cap. 2 sez. 2.5.2.1);
- (2) *tutelare un diritto* può essere analizzato anche come 'causare la continuazione dell'esistenza di un diritto', senso corrispondente alla funzione lessicale complessa PermFunc<sub>0</sub>;

La distinzione è sottile, si potrebbe addirittura contestare la validità della distinzione affermando che i due sensi siano solo due modi diversi di dire la stessa cosa. Ciò può forse essere vero nel caso in cui il soggetto di *tutelare* coincida col terzo attante semantico di DIRITTO<sub>2</sub>. Riteniamo tuttavia che la distinzione sia pertinente nei casi in cui il soggetto di *tutelare* è un'entità diversa. PermFunc<sub>0</sub> è una funzione lessicale complessa il cui senso è 'causare la continuazione dell'esistenza di...', o meglio 'agire in modo tale da non causare la fine di...' (Mel'čuk, 1996), dunque il soggetto del verbo può coincidere con qualunque agente che con la propria azione contribuisce a difendere l'esistenza di qualcosa. In diversi casi è stato difficile stabilire quale fosse il senso preciso del verbo *tutelare*, e dunque quale fosse la funzione lessicale corretta con cui etichettarlo. Il criterio di distinzione che si è deciso di utilizzare è la natura del soggetto del verbo *tutelare* nelle singole frasi:

- qualora il soggetto coincidesse con una legge o con un'autorità che incarna il potere legislativo, in altre parole chi ha il potere di stabilire un diritto (cioè il terzo attante semantico di DIRITTO<sub>2</sub>), abbiamo deciso di assegnare il tag corrispondente alla funzione lessicale Real<sub>3</sub>;
- qualora invece il soggetto fosse un'autorità o istituzione diversa, o altri soggetti che non hanno autorità legislativa (compresi il primo e il quarto attante di DIRITTO<sub>2</sub>) si è optato per la funzione lessicale PermFunc<sub>0</sub>.

Nei casi esaminati nei sottocorpora italiani di EPTIC il verbo *tutelare* compare in costruzioni impersonali (“la sicurezza è un diritto e in quanto tale deve essere tutelato in tutti i modi possibili”), oppure ha come soggetto generico l’Europa (“la cura e la protezione del minore sono divenuti diritti fondamentali che l’Europa tutela”), o una prima persona plurale generica che con ogni probabilità si riferisce al gruppo parlamentare di appartenenza dell’oratore (“noi vogliamo tutelare il diritto del consumatore a conoscere la provenienza di ogni prodotto”). Tutti i casi esaminati rientrano pertanto nella categoria definita dal secondo criterio sopra indicato, di conseguenza tutti sono stati etichettati con la funzione lessicale PermFunc<sub>0</sub>. Per chiarire la nostra scelta invitiamo il lettore a riflettere su un esempio non presente nel corpus, ideato da chi scrive: “La legge tutela il diritto allo studio”. In questo caso il soggetto è il terzo attante semantico di DIRITTO<sub>2</sub>, ed emerge più chiaramente il senso di realizzazione: la legge ‘fa ciò che deve fare’ (=Real<sub>3</sub>).

I casi dubbi, di cui abbiamo citato solo alcuni esempi, ci portano a riflettere sul fatto che tracciare distinzioni semantiche nette non sempre è facile. Un’annotazione di questo tipo, pur poggiando sulla categorizzazione formale proposta da Mel’čuk, che aspira ad essere rigorosa, pone dei problemi nel momento in cui si vanno ad interpretare le categorie semantiche per associarle a espressioni linguistiche concrete. Esiste il rischio che due annotatori indipendenti siano in disaccordo su come annotare casi ambigui. Proprio per questo motivo avvertiamo la necessità di uno studio semantico approfondito riguardante le funzioni lessicali e la loro applicazione nel trattamento delle collocazioni dell’italiano, che possa costruire un riferimento per il lavoro di annotazione. Per la lingua francese esistono numerosi studi di questo tipo, già citati nel presente lavoro; sarebbe auspicabile realizzare studi analoghi per la lingua italiana.

L’annotazione qui proposta presenta altri limiti. Innanzitutto essa riguarda un unico lessema, dunque le osservazioni che è possibile trarne sono limitate. Riteniamo tuttavia che, pur nelle sue dimensioni contenute, essa presenti caratteristiche interessanti. Ad esempio grazie all’annotazione per funzioni lessicali sono possibili ricerche mirate, in cui i collocati di un dato lessema sono filtrati in base al senso. Per questo motivo riteniamo possa essere uno strumento utile sia per l’apprendimento delle collocazioni, sia per la consultazione rapida ad esempio da parte di traduttori e interpreti. Segnaliamo inoltre che i collocati verbali di diritto individuati nei sottocorpora italiani di EPTIC e annotati nel presente studio di caso

corrispondono ai collocati verbali più salienti<sup>43</sup> riscontrati in un corpus italiano di lingua generale come itWaC:<sup>44</sup>

Collocato	Frequenza	Log Dice
Avere	135 649	8,086
Riconoscere	15 188	8,346
Esercitare	11 274	8,115
Garantire	10 518	7,832
Riservare	8 204	7,688
Tutelare	7 053	7,541
Difendere	5 730	7,195
Godere	4 886	6,989
Rispettare	4 310	6,761
Negare	4 285	6,813
Violare	4 273	6,853
Ottenere	4 076	6,396
Rivendicare	3 583	6,626

*Tabella 5: collocati verbali più frequenti del lessema DIRITTO nel corpus itWaC*

Ciò sembra indicare che nonostante EPTIC sia un corpus piccolo, i dati in esso contenuti possono comunque essere a loro modo significativi. Ricordiamo poi che lo studio di caso qui presentato serve soprattutto a far intuire le potenzialità di questo tipo di annotazione qualora si decidesse di estenderla alla totalità dei lessemi di un corpus (idealmente a un corpus annotato semanticamente, e di dimensioni superiori a EPTIC).

Un'altra possibile osservazione è che l'uso di corpora nell'apprendimento linguistico risulta difficoltoso per gli studenti, e che dunque l'annotazione qui proposta rischia di essere uno strumento che complica l'apprendimento anziché facilitarlo. A questa obiezione è possibile rispondere che l'uso dei corpora nella didattica delle lingue è consolidato da almeno un decennio, e che numerosi studi ne dimostrano l'efficacia (Bernardini e al. a cura di, 2003; Sinclair a cura di, 2004; Gallego-Hernandez e Rodriguez-Inés, 2016). I corpora possono inoltre costituire validi strumenti per l'apprendimento individuale, come sottolineato da Bernardini (2004) che ne incoraggia l'uso autonomo da parte degli studenti.

<sup>43</sup> Collocati verbali selezionati in base alla misura statistica Log Dice in un intervallo -3; +3.

<sup>44</sup> <https://www.sketchengine.co.uk/itwac-italian-corpus/>.

Infine, un'ulteriore possibile obiezione riguarda il fatto che la presente annotazione non possiede il carattere di "universalità" auspicato da Leech (2005), secondo il quale le annotazioni di corpora risultano tanto più utili e sfruttabili quanto più fanno riferimento a categorie linguistiche riconosciute, sulle quali esista un consenso diffuso. Le categorie linguistiche definite nella presente annotazione sono tutt'altro che universali, fanno anzi riferimento a una singola teoria non mainstream. Tuttavia, lo scopo globale del presente lavoro era precisamente introdurre la teoria Senso-Testo, di cui lo studio di caso rappresenta un'applicazione pratica. Ovviamente per poter sfruttare efficacemente le informazioni semantiche contenute nell'annotazione l'utente deve avere una minima conoscenza delle funzioni lessicali di Mel'čuk, ma confidiamo che con il supporto di un tagset dal taglio divulgativo, sul modello di quello riportato in sezione 4.4.1.1, l'utente possa sfruttare l'annotazione anche senza conoscere nel dettaglio la teoria delle funzioni lessicali. Ad ogni modo è opinione di chi scrive (basata sull'opinione di studiosi tra cui Mel'čuk (1998, 2015), Polguère (2003), L'Homme (2009, 2010) Primož (2011)) che l'introduzione delle funzioni lessicali nella didattica delle lingue risulterebbe estremamente utile, specialmente per futuri interpreti e traduttori, poiché "applied systematically and consistently within a given lexical field, the system of LFs can help translation students get a better grasp of the elusive collocability of lexemes" (Primož, 2011:129).

## Conclusioni

La tesi qui presentata costituisce un'introduzione alla teoria Senso-Testo e in particolare al formalismo delle funzioni lessicali, con lo scopo di illustrarne l'utilità nel trattamento delle collocazioni linguistiche.

Col presente lavoro abbiamo voluto dimostrare, anche attraverso un'applicazione computazionale quale l'annotazione di corpora, che le funzioni lessicali permettono di organizzare in modo sistematico, in base a criteri semantici e sintattici, le collocazioni di una lingua e che potrebbero dunque facilitarne l'apprendimento.

Il capitolo 1 è stato dedicato alla presentazione generale della teoria Senso-Testo all'interno della quale le funzioni lessicali sono state concepite. Il capitolo 2 ha introdotto il concetto di collocazione linguistica e presenta nel dettaglio le funzioni lessicali, lo strumento offerto della teoria Senso-Testo per descrivere e rappresentare formalmente le collocazioni. Nel capitolo 3 è stato fornito un background sull'annotazione di corpora, introducendo innanzitutto il concetto di corpus e illustrandone le tipologie, e dedicando poi uno spazio alla teoria e pratica dell'annotazione (oggetto dello studio di caso del capitolo 4). Infine, il capitolo 4 è stato dedicato ad uno studio di caso con il quale si è voluto proporre un'annotazione per funzioni lessicali del corpus EPTIC, che idealmente, e qualora sia sviluppata andando a coprire la totalità delle collocazioni presenti nel corpus, potrebbe costituire un utilissimo strumento per l'apprendimento delle collocazioni, destinato in particolare agli studenti di traduzione e interpretazione non italofoeni.

Con il presente lavoro speriamo di aver contribuito a stimolare l'interesse verso la teoria Senso-Testo e in particolare verso le funzioni lessicali. Ci auguriamo che in futuro questa teoria attualmente poco studiata in Italia possa avere maggiore spazio. Esistono numerosi ambiti che sarebbe interessante esplorare, ad esempio lo sviluppo di strumenti lessicografici per la lingua italiana analoghi a quelli esistenti per altre lingue: un dizionario esplicativo-combinatorio italiano, un database lessicale analogo al DiCouèbe francese, o un database di collocazioni dell'italiano sul modello del già citato DiCoLiLex (cfr. cap. 4 sez. 4.4.1.1). Un altro ambito potenzialmente interessante è l'utilizzo delle funzioni lessicali in didattica delle lingue. Numerosi studiosi hanno infatti indicato i potenziali benefici dell'introduzione delle funzioni

lessicali nei curricula universitari dedicati alle lingue e alla traduzione: sarebbe estremamente interessante condurre studi sperimentali sui risultati ottenuti in termini di apprendimento.

## Bibliografia

- Agel, V., Eichinger, L.M., Eroms, H-W., Hellwig, P., Heringer, H.J., Lobin, H. a cura di (2003). *Dependency and Valency. An International Handbook of Contemporary Research*. New York: De Gruyter.
- Aijmer, K. e Altenberg, B. a cura di (1991). *English Corpus Linguistics*. London: Longman.
- Alonso Ramos, M. e Tutin, A. (1996). "A Classification and Description of Lexical Functions for the Analysis of their Combinations". In Wanner L. a cura di (1996), 147-167.
- Anctil, D. (2005) *Maîtrise du lexique chez les étudiants universitaires: typologie des problèmes lexicaux et analyse des stratégies de résolution de problèmes lexicaux*. Tesi di laurea magistrale. Université de Montréal [consultata il 27.02.2018]
- Andor, J. (2004). "The master and its performance. An interview with Noam Chomsky". *Intercultural pragmatics*, 1: 93-111.
- Anthony, L. (2005). "AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom". In *IEEE International Professional Communication Conference*, 729-737.
- Bazell, C. E., Catford, J. C., Halliday, M. A. K. e Robins, R. H. a cura di (1966). *In memory of J. R. Firth*. London: Longman.
- Bernardini, S. (2004). "Corpora in the classroom. An overview and some reflections on future developments". In Sinclair J. a cura di (2004), 15-36.
- Bernardini, S., Ferraresi, A. e Miličević, M. (2016). "From EPIC to EPTIC – Exploring simplification in interpreting and translation from an intermodal perspective". *Target*, 28(1): 58-83.
- Bernardini, S., Stewart, D., e Zanettin, F. a cura di (2003). *Corpora in translator education*. Manchester: St. Jerome.
- Biber, D. (1993). "Using register-diversified corpora for general language studies". *Computational Linguistics*, 19: 219-241.
- Boas, F. (1940). *Race, Language and Culture*. New York: Macmillan.
- Chomsky, N. (1970). *Saggi linguistici*. Torino: Boringhieri.
- Clark, A., Fox, C. e Lappin, S. a cura di (2010). *The Handbook of Computational Linguistics and Natural Language Processing*. Malden: Blackwell Handbooks in Linguistics.
- Comrie, B. e Polinsky, M. a cura di (1993). *Causativity and Transitivity*. Amsterdam/Philadelphia: John Benjamins.
- Cowie, A. P. a cura di (1998). *Phraseology: theory, analysis and applications*. New York: Oxford University Press.

- De Mauro, T. (2016) *Nuovo vocabolario di base della lingua italiana*. disponibile online all'indirizzo <https://dizionario.internazionale.it/> [consultato in data 06.02.2018]
- Dostie 1999: "Méthodologie d'élaboration des articles du Dictionnaire explicatif et combinatoire du français contemporain". In Mel'čuk e al. (1984-1999), vol 4, 11-28.
- Fellbaum, C. a cura di (1998). *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.
- Ferraresi, A. e Bernardini, S. (in stampa) "Building EPTIC: A many-sided, multi-purpose corpus of EU Parliament proceedings". In M. Sánchez Nieto and I. Doval (eds.) *Parallel Corpora: Creation and Application*. Amsterdam/Philadelphia: John Benjamins.
- Fontenelle, T. (1992). "Collocation acquisition from a corpus or from a dictionary: a comparison". In Tommola, Varantola, Salmi-Tolonen e Schopp a cura di (1992). *EURALEX 92 Proceedings I-I*, Fifth EURALEX International Congress, Studia Translatologica, Ser. A, Vol.1, University of Tampere, 220-228.
- Fontenelle, T. (1998). "Discovering significant lexical functions in dictionary entries". In Cowie A. P. (1998), 189-207.
- Frassi, P. Tremblay, O. (2016). "Il Réseau Lexical du Français: una banca dati per l'apprendimento del lessico francese". In Bianchi, F., Leone, P. (2016). *Linguaggio e apprendimento linguistico. Metodi e strumenti tecnologici*. Milano: AiTLA, 155-172.
- Gagné, A. M. e L'Homme M. C. (2016). "Opposite relationships in terminology". *Terminology*, 22(1):30-51.
- Gallego-Hernandez, D.e Rodriguez-Inés, P. a cura di (2016). "Special Issue: Corpus Use and Learning to Translate, almost 20 years on". *Cadernos de Tradução*, 36(1).
- Gatti, M.C. (1992). *Dalla semantica alla Lessicologia. Introduzione al modello Senso-Testo di I. A. Mel'čuk*. Brescia: La Scuola.
- Gelbukh, A. a cura di (2002). "Computational Linguistics and Intelligent Text Processing". *Proceedings of the CICLing 2002 Mexico City. Lecture Notes in Computer Science*, vol 2276. Berlin, Heidelberg: Springer.
- Granger, S. (1998). "Prefabricated patterns in advanced EFL writing: collocations and formulae" In Cowie (1998), 145-160.
- Granger, S. e Meunier, F. a cura di (2008). *Phraseology: An Interdisciplinary Perspective*. Amsterdam/Philadelphia: John Benjamins.
- Granger, S. e Paquot, M. a cura di (2012). *Electronic Lexicography*. Oxford: Oxford University Press.
- Greenbaum, S. (1970). *Verb-Intensifier Collocations in English. An Experimental Approach*. The Hague: Mouton.
- Gries, S. T. (2008). "Phraseology and Linguistic Theory. A Brief Survey". In Granger S. e Meunier F. a cura di (2008), 3-26.

- Grossmann, F. e Tutin, A. a cura di (2003). "Les Collocations. Analyse et traitement". *Travaux et Recherches en Linguistique Appliquée*, E:1. Amsterdam: De Werelt.
- Halliday, M. A. K. (1966). "Lexis as a linguistic level". In Bazell C. e al. a cura di (1966), 148-162.
- Harris, R. A. (1993). *The Linguistic Wars*. Oxford: Oxford University press.
- Howarth, P. (1996). *Phraseology in English Academic Writing. Some Implications for Language Learning and Dictionary Making*. Tübingen: Niemeyer
- Howarth, P. (1998). "The phraseology of learners' academic writing". In Cowie, A. P. (1998). 161-186.
- Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge: MIT Press.
- Jakobson, R. (1960). "Closing Statements: Linguistics and Poetics". In Sebeok, T.A. (1960). *Style In Language*, Cambridge: MIT Press, 350-377.
- Jones, S. e Sinclair, J. (1974). "English lexical collocations". *Cahiers de lexicologie*, 24: 15-61.
- Jousse, A. L. (2007). "Formalizing non-standard lexical functions". *Wiener Slawistischer Almanach*, Sonderband 69: 219-228.
- Jousse, A. L. (2010). *Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales*. Tesi di dottorato. Université de Montréal. <http://olst.ling.umontreal.ca/pdf/TheseALJousse.pdf> [consultata il 05.05.2017]
- Jousse, A.L. e Polguere, A. (2005). *Le DiCo et sa version DiCouèbe - Document descriptif et manuel d'utilisation*. Observatoire de linguistique Sens-Texte, Université de Montréal [documento in formato solo elettronico consultato in data 22.11.2017 all'indirizzo <http://idefix.ling.umontreal.ca/dicouebe/DiCoDOC.pdf> ]
- Jurafsky, D. e Martin, J.H. (2008). *Speech and Language Processing*. London: Prentice Hall.
- Kahane, S. (2003). "The Meaning-Text Theory". In Agel e al. a cura di (2003). Vol. 1, 546-570.
- Kahane, S. e Polguère, A. (2001). "Formal Foundations of Lexical Functions". In Daille, B. e Williams, G. a cura di (2001) *Proceedings of the ACL Workshop on Collocation: Computational Extraction, Analysis and Exploitation*, Toulouse, 8-15.
- Kilgarriff, A. e Kosem, I. (2012). "Corpus Tools for Lexicographers". In Granger S. e Paquot M. a cura di (2012), 31-55.
- Kjellmer, G. (1994). *A Dictionary of English Collocations: Based on the Brown Corpus*. 3 volumes. Oxford: Clarendon Press.
- L'Homme, M.C. (2009). "A Methodology for Describing Collocations in a Specialized Dictionary". Nielsen, S. and S. Tarp (eds.). *Lexicography in the 21st Century In honour of Henning Bergenholtz*. Amsterdam/Philadelphia: John Benjamins.

- L'Homme, M.C. (2010). "Designing Terminological Dictionaries for Learners based on Lexical Semantics: The representation of actants". In Fuertes-Olivera, P. (ed.). *Specialised Dictionaries for Learners*, Berlin/New York: De Gruyter, pp. 141-153.
- Lambrey, F. (2016). *Implémentation des collocations pour la réalisation de texte multilingue*. Tesi di Laurea Magistrale. Université de Montréal [http://olst.ling.umontreal.ca/pdf/Lambrey Florie 2016 memoire.pdf](http://olst.ling.umontreal.ca/pdf/Lambrey_Florie_2016_memoire.pdf) [consultata il 05.04.2017].
- Leech, G. (1991). "The state of the art in corpus linguistics". In Aijmer K. e Altenberg B. a cura di (1991), 8-29.
- Leech, G. (2005). "Adding Linguistic Annotation". In Wynne M. a cura di (2005), 17-29.
- Lenci, A., Montemagni, S. e Pirrelli, V. (2005). *Testo e computer*. Roma: Carocci
- Marengo, S. e Robichaud B. (approvato). "Une nouvelle ressource pour la description des collocations: le DicoLiLex, dictionnaire de liens lexicaux", *Cahiers de Lexicologie*. Riassunto consultabile al link [http://rsl7.evenement.usherbrooke.ca/resumes\\_communications/marengo\\_et%20al resum e\\_RSL7.pdf](http://rsl7.evenement.usherbrooke.ca/resumes_communications/marengo_et%20al_resum_e_RSL7.pdf)
- McEnery, T. e Wilson, A. (2001). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Mel'čuk, I. A. (1974). *Опыт теории лингвистических моделей Смысл-Текст* [Opyt teorii lingvističeskich modelej Smysl-Tekst. It: Verso una teoria dei modelli linguistici Senso-Testo]. Mosca: Nauka. Ristampato (1999) Mosca: Jazyki ruskoj kul'tury.
- Mel'čuk, I. A. (1976). *Das Wort: zwischen Inhalt und Ausdruck*. München: Fink.
- Mel'čuk, I. A. (1996). "Lexical functions: a tool for the description of lexical relations in a lexicon". In Wanner L. a cura di (1996), 37-102.
- Mel'čuk, I. A. (1998). "Collocations and Lexical Functions", in Cowie P. a cura di (1998), 23-53.
- Mel'čuk, I. A. (2003). "Les collocations: définition, rôle et utilité", in Grossmann F. e Tutin A. a cura di (2003), 23-31.
- Mel'čuk, I. A. (2012). *Semantics: from meaning to text (Vol.I)*. Amsterdam/Philadelphia: John Benjamins.
- Mel'čuk, I. A. (2013). *Semantics: from meaning to text (Vol.II)*. Amsterdam/Philadelphia: John Benjamins.
- Mel'čuk, I. A. (2013b). "Tout ce que nous voulions savoir sur les phrasèmes, mais...". *Cahiers de lexicologie*, 102: 129-150.
- Mel'čuk, I. A. (2015). "Clichés, an Understudied Subclass of Phrasemes". *Yearbook of Phraseology*. 6: 55-86.
- Mel'čuk, I. A. e Polguère, A. (2007). *Lexique actif du français: L'apprentissage du vocabulaire fondé sur 20000 dérivations sémantiques et collocations du français*. Paris: Duculot.

- Mel'čuk, I. A., Clas, A. e Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Bruxelles: Duculot.
- Mel'čuk, I., Arbatchewsky-Jumarie, N., Iordanskaja, L., Mantha, S. e Polguère, A., (1984, 1988, 1992, 1999). *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexicosémantiques I, II, III, IV*. Montréal: Presses de l'Université de Montréal
- Milicevic, J. (2003). *Modélisation sémantique, syntaxique et lexicale de la paraphrase*. Tesi di dottorato. Université de Montréal. [consultata il 20.10.2017]
- Miličević, J. (2006). "A short guide to the Meaning-Text linguistic theory". *Journal of Koralex*, 8: 187-233. <http://olst.ling.umontreal.ca/pdf/IntroMTTJM.pdf>
- Nattinger, J.R. e DeCarrico, J. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nérima, L., Seretan, V., e Wehrli, E. (2006). "Le problème des collocations en TAL". *Nouveaux cahiers de linguistique française*, 27: 95-115.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam/Philadelphia: John Benjamins.
- Papp, F. (1966). *Mathematical Linguistics in the Soviet Union*. The Hague: Mouton.
- Polguère, A. (1998). "La théorie Sens-Texte". *Dialangue*, Vol. 8-9: 9-30.
- Polguère, A. (2000). "Une base de données lexicales du français et ses applications possibles en didactique". *Revue de linguistique et de didactique des langues (Lidil)*, 21: 75-97
- Polguère, A. (2003). "Collocations et fonctions lexicales: pour un modèle d'apprentissage", in Grossmann F. e Tutin A. a cura di (2003), 117-133.
- Polguère, A. (2011). "Classification sémantique des lexies fondée sur le paraphrasage". *Cahiers de lexicologie*, 98: 197-211.
- Polguère, A. (2012). "Propriétés sémantiques et combinatoires des quasi-prédicats sémantiques". *Scolia*, 26: 131-152.
- Polguère, A. e Mel'čuk, I. A. (2006) "Dérivations sémantiques et collocations dans le DiCo/LAF". *Langue française*, 150: 66-83. (consultato nella sua versione online all'indirizzo [www.persee.fr/doc/lfr\\_0023-8368\\_2006\\_num\\_150\\_2\\_6854](http://www.persee.fr/doc/lfr_0023-8368_2006_num_150_2_6854) ).
- Primož, J. (2011). "Meaning-Text Theory in the translator's classroom". *Rivista internazionale di tecnica della traduzione - International Journal of Translation*, 13:129-138.
- Prencipe, V. (2006). *Traduzione come doppia comunicazione: un modello senso-testo per una teoria linguistica della traduzione*. Milano: Franco Angeli.
- Rigotti, E. (1973). "La linguistica in Russia dagli inizi del secolo XIX ad oggi. IV. Il ritorno dello strutturalismo in URSS e i suoi sviluppi". *Rivista di Filosofia neoscolastica*, 3: 488-521.
- Rigotti, E. (1979). *Principi di teoria linguistica*. Brescia: La Scuola.

- Rychlý, P. (2007). "Manatee/Bonito: A Modular Corpus Manager". *1st Workshop on Recent Advances in Slavonic Natural Language Processing*, 65–70. Masaryk University, Brno.
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A. e Flickinger, D. (2002). "Multiword Expressions: A Pain in the Neck for NLP". In Gelbukh a cura di (2002), 1-15.
- Samson-Fauteux, M. (2010). *Représentations du sens lexical en lexicologie explicative et combinatoire: étude de trois formalisations spécifiques*. Tesi di Laurea Magistrale. Université de Montréal [https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/4554/Samson-Fauteux\\_Melissa\\_2010\\_memoire.pdf?sequence=2](https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/4554/Samson-Fauteux_Melissa_2010_memoire.pdf?sequence=2) [consultata il 05.09.2017].
- Sandrelli, A. e Bendazzoli, C. (2005). "Lexical Patterns in simultaneous interpreting: a preliminary investigation of EPIC (European Parliament Interpreting Corpus)". *Proceedings from the Corpus Linguistics Conference Series 1/1*. Birmingham: University of Birmingham. 1-19.
- Santandrea, M. (2014). *Le collocazioni in traduzione e interpretazione tra italiano e inglese: uno studio su EPTIC\_01\_2011*. Tesi di laurea magistrale. Università degli Studi di Bologna. [consultata il 21.01.2018].
- Saussure, F. (1916) *Cours de linguistique générale*. Consultato nell'edizione critica a cura di Tullio de Mauro (1967) Paris: Payot.
- Shlesinger, M. (2009). "Towards a definition of Interpretese: an intermodal, corpus-based study". In Hansen, G., Chesterman, A. e Gerzymisch-Arbogast, H. (2009). *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile*. Amsterdam: John Benjamins, 237-253.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2005). "Corpus and Text – Basic Principles". In Wynne M. a cura di (2005), 1-16.
- Sinclair, J. a cura di (2004). *How to Use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins.
- Smadja, F. (1991) "Macro-coding the lexicon with co-occurrence knowledge". in Zernik a cura di (1991) *Lexical Acquisition: Using On-Line Resources to Build a Lexicon*. Hillsdale: LEA, 165-189.
- Sosoni, V. (2011). "Training translators to work for the EU institutions: luxury or necessity?". *Journal of Specialized Translation*, 16: 77-108.
- Stragapede, E. (2017). *Allargamento e analisi di un corpus intermodale: le parole chiave del sottocorpus di italiano di EPTIC\_2011*. Tesi di laurea magistrale. Università degli Studi di Bologna. [consultata il 21.01.2018]
- Tamburini, F. (2005). "Elaborazione automatica del linguaggio parlato". In Rossini Favretti R. a cura di (2005), *E-Learning: comunicazione mediata e presenziale*. Bologna: BUP, 39-54.
- Tiberii, P. (2012). *Dizionario delle collocazioni: le combinazioni delle parole in italiano*. Bologna: Zanichelli.

- Tremblay, O. (2009). Une ontologie des savoirs lexicologiques pour l'élaboration d'un module de cours en didactique du lexique. Tesi di dottorato. Université de Montréal. <http://olst.ling.umontreal.ca/pdf/OphelieTremblayThese2009.pdf> [consultata in data 23.01.2018].
- Tutin, A. (2005). "Le dictionnaire de collocations est-il indispensable?". *Revue française de linguistique appliquée*, X (2): 31-48.
- Tutin, A. (2008). "For an extended definition of lexical collocations". *Proceedings of the 13th EURALEX International Congress*, Barcelona, 1453-1460. <https://halshs.archives-ouvertes.fr/halshs-00371418>.
- Urzi, F. (2009). *Dizionario delle combinazioni lessicali*. Luxembourg: Convivium.
- Versace, M. (2011). *La Russkaja jazykovaja kartina mira: teorie, metodi di analisi e applicazioni*. Tesi di dottorato. Università Cattolica del Sacro Cuore, Milano.
- Wanner, L. a cura di (1996). *Lexical functions in Lexicography and natural language processing*. Amsterdam/Philadelphia: John Benjamins.
- Wynne, M. a cura di (2005). *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbrow Books.
- Žolkovskij, A. K. e Mel'čuk, I. A. (1965). "O vozmožnom metode i instrumentach semantičeskogo sinteza" [Su un possibile metodo e sugli strumenti di sintesi semantica]. *Naučno-techničeskaja informacija*, 6: 23-28.
- Žolkovskij, A. K. e Mel'čuk, I. A. (1966). "O sisteme semantičeskogo sinteza. I. Stroenie slovarja" [Su un sistema per la sintesi semantica. 1. Struttura del dizionario]. *Naučno-techničeskaja informacija*, 11: 48-55.
- Žolkovskij, A. K. e Mel'čuk, I. A. (1967). "O semantičeskom sintese" [Sulla sintesi semantica]. *Problemy kibernetiki*, vol. 19, 177-238. Traduzione inglese pubblicata nel 1970 in *Systems Theory Research*, 19: 179-246.



## Appendice 1: Glossario

Il presente glossario raccoglie le definizioni dei principali concetti legati alla teoria e al modello Senso-Testo. È stato pensato sia come supporto alla lettura dei capitoli 1 e 2, di cui riprende le nozioni-chiave, sia come strumento per la consultazione indipendente, destinato a chi, ad esempio, si approccia per la prima volta alla teoria Senso-Testo e desidera avere un quadro completo ma sintetico dei principali concetti.

Nel formulare le definizioni ci si è basati su spiegazioni tratte dalla letteratura Senso-Testo (citando gli autori e le opere di riferimento), che sono state rielaborate in modo da renderle più esplicite e chiare anche per chi non ha familiarità con la teoria Senso-Testo.

Gli esempi in lingua italiana sono stati ideati da chi scrive. Sono presenti anche esempi in lingua inglese tratti dalla letteratura.

Per ciascun termine sono indicate sia le sezioni del presente lavoro in cui il termine è citato e trattato per esteso, sia i riferimenti bibliografici a cui si rimanda per approfondimenti.

attante	<p>Nella teoria Senso-Testo un attante è un <i>argomento</i>.</p> <p>Gli <b>attanti semantici</b> sono gli argomenti (obbligatori) che un predicato (o un quasi-predicato) semantico possiede, ossia i partecipanti essenziali alla situazione indicata dal predicato (Mel'čuk, 2012). Ad esempio, il predicato semantico <i>omicidio</i> ha sostanzialmente due attanti: <i>l'omicida</i> (X) e <i>la vittima</i> (Y) e. Il quasi-predicato <i>sorella</i> ha due argomenti: <i>la sorella stessa</i> (X) e <i>la persona di cui X è sorella</i> (Y). All'interno della struttura semantica gli attanti sono numerati e indicati con numeri arabi posti in corrispondenza degli archi che congiungono un elemento ai suoi attanti. (Cfr. cap.1 sez. 1.5.2.1.1).</p> <p>Gli <b>attanti sintattici profondi</b> sono gli argomenti sintattici di un verbo o di un'altra unità lessicale (anche alcuni sostantivi possono avere degli attanti). All'interno della struttura sintattica profonda gli attanti sono indicati con numeri romani posti in corrispondenza dei rami del grafo ad albero (cfr. cap. 1 sez. 1.5.2.2).</p>
collocazione	<p>Nella linguistica Senso-Testo, frasema lessicale compositiva, espressione formata da più elementi la cui composizione è il frutto di restrizioni lessicali precise: uno degli elementi (detto <i>base</i> della collocazione) è scelto liberamente dal locutore per esprimere un senso, mentre la scelta degli altri elementi (i <i>collocati</i> della base) è condizionata, oltre che in funzione del senso da esprimere, dalle <i>proprietà combinatorie</i> del primo elemento (Mel'čuk 2013b) (cfr. cap.2 sez. 2.3.2).</p>
definizione	<p>La definizione di un elemento linguistico (un lessema, una locuzione o un altro tipo di espressione linguistica) è l'esplicitazione del suo contenuto semantico, del suo senso. Il senso di un'espressione non è</p>

direttamente accessibile (cfr. **sensio**<sup>α</sup>), dunque per esplicitare il senso di un'espressione in genere si fa ricorso a una parafrasi di tale espressione, ossia a un'altra espressione che abbia quasi lo stesso senso: questo è sia il modo in cui sono costruite le definizioni lessicografiche, sia il modo in cui i parlanti procedono intuitivamente quando sono chiamati a dare la definizione di un termine (Polguère, 2003).

In lessicografia Senso-Testo (cfr. cap. 1 sez. 1.6.1.1) definizioni sono basate sul modello aristotelico di *definizione per genere prossimo e differenza specifica*, ossia una definizione che "individua una più vasta classe di enti (genere prossimo) entro la quale si trova la sottoclasse degli oggetti da definire, rilevando poi la o le restanti caratteristiche peculiari dell'oggetto che si vuol definire (differenza specifica): per es., «l'uomo è un animale [genere prossimo] razionale [differenza specifica]» (Vocabolario Treccani). Per approfondimenti sulle modalità di elaborazione delle definizioni nella lessicografia esplicativa-combinatoria si veda Dostie (1999).

derivazione semantica      Operazione che permette di ottenere, a partire da un lessema  $L_1$ , un lessema  $L_2$  il cui senso corrisponde a uno degli attanti semantici di  $L_1$ , cioè uno dei partecipanti alla situazione designata da  $L_1$ ;  $L_2$  è allora detto *derivato semantico* di  $L_1$  (Mel'čuk, 1996).

Esempio: *omicidio* ( $L_1$ ) → *vittima* ( $L_2$ ).

Nella teoria Senso-Testo, la derivazione semantica designa più specificamente il **tipo di relazione lessicale** esistente tra un lessema  $L_1$  e un lessema  $L_2$  che corrisponde a un attante semantico di  $L_1$  (o meglio, un lessema che costituisce la *realizzazione lessicale* di un attante semantico di  $L_1$ ).

La relazione lessicale che lega il lessema *omicidio* al lessema *vittima* è dunque definita, nella teoria Senso-Testo, una relazione di derivazione semantica.

Le funzioni lessicali che rappresentano i vari modi in cui la relazione di derivazione semantica si esprime nel lessico di una lingua sono:  $S_i$ ,  $A_i$ ,  $Adv_i$  (cfr. Appendice 2: Funzioni Lessicali Standard Semplici).

Altre funzioni lessicali che rappresentano relazioni di derivazione lessicale sono  $S_{instr}$ ,  $S_{med}$ ,  $S_{mod}$ ,  $S_{loc}$ ,  $S_{res}$ . Tuttavia, in questi ultimi casi la relazione tra  $L_1$  e  $L_2$  è leggermente diversa, perché qui  $L_2$  non indica non un attante semantico di  $L_1$  bensì un *circostante*, un elemento facoltativo legato alla situazione semantica di  $L_1$  (cfr. Appendice 2).

derivazione sintattica      **Operazione** che permette di ottenere, a partire da un lessema  $L_1$ , un lessema  $L_2$  dotato dello stesso significato, ma appartenente a una categoria sintattica diversa;  $L_2$  è allora detto *derivato sintattico* di  $L_1$ .

Esempio: *fraintendere* ( $L_1$ ) → *fraintendimento* ( $L_2$ ).

Mel'čuk (2012:103) offre la seguente definizione formale di derivato sintattico: "Un lessema  $L_2$  è un derivato sintattico di un lessema  $L_1$  se e solo se:

- 1)  $L_1$  e  $L_2$  non appartengono alla stessa categoria sintattica
- 2) il loro contenuto semantico è sostanzialmente lo stesso
- 3) i loro attanti sintattici profondi corrispondono ai loro attanti semantici".

Il concetto di derivazione non presuppone l'esistenza di una categoria grammaticale "originaria", da cui le altre categorie discendono; in altre parole, è impossibile stabilire se un sostantivo sia "nato" dal verbo corrispondente o viceversa. La derivazione è un processo dinamico che agisce in tutte le direzioni, poiché è (quasi) sempre possibile, ad esempio, passare da un sostantivo all'aggettivo corrispondente, o al verbo, senza che questo passaggio presupponga una gerarchia.

Oltre a indicare l'operazione sopra descritta, nella teoria Senso-Testo il termine derivazione sintattica designa più nello specifico il **tipo di relazione lessicale** che unisce un lessema a un suo derivato sintattico. Utilizzando l'esempio sopra citato, la relazione lessicale che lega *fraintendere* a *fraintendimento* è la relazione di derivazione sintattica.

Le funzioni lessicali che rappresentano questo tipo di relazione lessicale sono  $V_0$ ,  $S_0$ ,  $A_0$ ,  $Adv_0$  (cfr. Appendice 2: Funzioni Lessicali Standard Semplici).

frasema

Categoria lessicale che comprende le espressioni formate da più elementi, frutto di una combinazione non libera: un frasema è un'espressione caratterizzata da un certo grado di fissità. All'interno della categoria dei frasemi è possibile individuare diverse sottocategorie: frasemi semantici, frasemi pragmatici, frasemi lessicali (Mel'čuk, 2013b).

A seconda del grado di fissità e di composizionalità i frasemi (lessicali) possono essere ulteriormente distinti in locuzioni idiomatiche e le collocazioni (Cfr. cap. 2 sez. 2.3.2).

funzione lessicale

Strumento formale che permette di rappresentare e descrivere in maniera compatta e sistematica le relazioni lessicali esistenti tra le parole di una data lingua, e che di conseguenza permette di rappresentare le collocazioni linguistiche (Mel'čuk 1996). (Cfr. cap. 2 sez. 2.4.1).

Il termine funzione è impiegato nel suo senso matematico  $f(x) = y$ , poiché una funzione lessicale associa a una data espressione  $L$  (che costituisce l'argomento, o base, della funzione  $f$ ) un insieme  $\{L_i\}$  di espressioni (ossia i valori della funzione  $f$ ) che esprimono, in relazione a  $L$ , un significato specifico associato a  $f$ :  $f(L) = \{L_i\}$ .

Si considerino le seguenti collocazioni della lingua italiana: “stanco morto”; “povero in canna”; “ricco sfondato”. Ciò che accomuna queste espressioni è la relazione tra il primo e il secondo elemento: si tratta di una relazione di senso che possiamo genericamente definire ‘intensità’. Tale relazione può essere rappresentata a livello formale mediante la funzione lessicale Magn: Magn(stanco) = morto; Magn(povero) = in canna; Magn(ricco) = sfondato.

Magn è solo un esempio. Esistono una settantina di funzioni lessicali standard semplici (cfr. Appendice 2: Funzioni Lessicali Standard Semplici), ciascuna identificata da un nome che è l’abbreviazione di una parola latina o greca e che funge da etichetta esplicativa.

Nell’architettura del modello Senso-Testo le funzioni lessicali si collocano nel livello sintattico profondo (vedi capitolo 1 sez. 1.5.2.2): esse compaiono tra le **unità lessicale**<sup>α</sup> profonde con le quali vengono contrassegnati i nodi dell’albero sintattico profondo (Gatti, 1992:65), assieme ai **lessema**<sup>α</sup> e alle **locuzione**<sup>α</sup> (Milićević, 2006).

Le funzioni lessicali hanno un ruolo essenziale all’interno del modello Senso-Testo e anche all’interno del Dizionario esplicativo-combinatorio (cfr. cap. 1 sez. 1.6), poiché esse descrivono e rappresentano le proprietà combinatorie di ciascun lessema. Nel Dizionario esplicativo-combinatorio per ciascuna voce lessicale si enumerano le relative funzioni lessicali, che vanno a collocarsi in una zona apposita prevista per la descrizione sistematica delle proprietà combinatorie dei lessemi” (cfr. cap. 1 sez. 1.6.1.3).

lessema

Categoria lessicale: insieme di parole-forma che si distinguono unicamente per la flessione (Mel’čuk, 2012). Un lessema possiede un (singolo) significato lessicale che è condiviso da tutte le parole-forma di tale lessema: le parole-forma *bambino*, *bambina*, *bambini*, *bambine* condividono il medesimo significato (= essere umano di età compresa tra la nascita e l’adolescenza), e appartengono al lessema *bambino*, ma si distinguono per genere e numero.

Concretamente, un lessema è la “forma base” di una parola, cioè quella che tipicamente costituisce l’entrata di una voce in un dizionario.

Nel modello Senso-Testo i lessemi costituiscono i nodi della struttura sintattica profonda, assieme alle *locuzioni* e alle *funzioni lessicali* (cfr. cap. 1 sez. 1.5.2.2).

Il termine lessema è proprio dell’approccio Senso-Testo, ma presenta significative analogie con la nozione linguistica di *lemma*.

locuzione idiomatica  
(o semplicemente  
locuzione)

Frasema lessicale non-composizionale (Mel’čuk, 2013b), espressione fissa che non è frutto di composizione – cioè il cui significato non è costruito a partire dalla somma dei significati delle parole che la compongono – bensì è dotata di un significato proprio e specifico (Cfr. cap. 1 sez. 1.5.2.2 e cap. 2 sez. 2.3.2.).

Una locuzione funziona come un unico blocco: innanzi tutto presenta una forte coesione sintattica, cioè non è quasi mai possibile separare o modificare le parole di cui è composta (ad esempio: dormire sugli allori → \*Marco ha dormito su *molti* allori); inoltre, una locuzione presenta alcune importanti caratteristiche in comune con i lessemi, ossia il fatto di rappresentare una forma-base a cui corrispondono varie forme flesse (dormire sugli allori → *tu dormi sugli allori, ho dormito sugli allori, non dormite sugli allori!* etc.). La differenza rispetto ai lessemi è che le locuzioni idiomatiche sono necessariamente formate da più elementi, mentre in genere i lessemi sono espressioni formate da un solo elemento.

modello funzionale

Per modello funzionale Y del fenomeno X oggetto di studio si intende un sistema artificiale il cui comportamento è sufficientemente simile a quello di X, tale che Y, a partire da uno stesso input, genera lo stesso output generato da X (Mel'čuk, 2012). I modelli funzionali sono comunemente utilizzati in altre scienze (ad esempio in fisica) per descrivere e rappresentare fenomeni non direttamente osservabili.

Come sottolinea lo stesso Mel'čuk (citato in Gatti, 1992:30), "poiché le medesime conseguenze possono derivare da premesse diverse, l'identità di comportamento tra X e Y non garantisce una loro identità strutturale". Per questo motivo il modello Senso-Testo in questione è detto *funzionale* e non *strutturale* (Milićević, 2006). (Cfr. cap. 1 sez. 1.4.1).

nome semantico

Semantema che non implica nessun attante, e che in generale denota un'entità, come ad esempio nomi di individui, luoghi, date, sostanze (Mel'čuk, 2012). (Cfr. cap.1 sez. 1.5.2.1.1).

parafrasi

Un'espressione F' è una parafrasi di un'altra espressione F se i loro significati coincidono, ossia se il loro contenuto semantico (il loro senso linguistico) è lo stesso (si noti l'analogia con la nozione di sinonimia). Così come è molto difficile riscontrare una sinonimia perfetta (equivalenza perfetta tra il senso di due termini diversi) nelle lingue naturali, una parafrasi F' avrà quasi sempre un senso leggermente diverso rispetto a F.

La nozione di parafrasi è centrale nella definizione di **senso** offerta dalla teoria Senso-Testo (cfr. cap. 1 sez. 1.4.1 e 1.5.2.1). Inoltre, i meccanismi di parafrasi svolgono un ruolo importante all'interno del modello Senso-Testo. Per approfondimenti sulla parafrasi e sul ruolo dei meccanismi parafrastici nella modello Senso-Testo si rimanda a Mel'čuk (2012) e a Milićević (2003).

*parola-forma*<sup>45</sup>

Segno linguistico segmentale autonomo (che compare cioè tra due spazi) e minimo (cioè che non può essere scomposto in altre parole-forma) (Mel'čuk, 2012). Una parola-forma può tuttavia essere scomposta in segni più semplici, ad esempio nei morfemi che la caratterizzano.

Concretamente, una parola-forma corrisponde a una forma flessa di un lessema: *vado, andiamo, andavate, andresti* sono tutte parole-forma, il cui lessema di riferimento è il verbo *andare*.

predicato semantico

A livello di rappresentazione semantica esistono due classi principali di semantemi (cfr. semantema): i nomi semantici e i predicati.

Sono considerati predicati tutti i semantemi che denotano situazioni che implicano *partecipanti* (detti *attanti* → cfr. *attente*), cioè situazioni in cui più elementi (di varia natura) sono posti in relazione tra loro. Tipicamente, i predicati denotano eventi, azioni, processi, stati, percezioni, relazioni, proprietà, quantità, eccetera (Mel'čuk, 2012).

I predicati semantici non si limitano ai verbi: anche molti sostantivi e avverbi, a livello semantico, costituiscono dei predicati. (Cfr. cap.1 sez. 1.5.2.1.1).

quasi-predicato  
(semantico)

Semantema che, pur denotando un'entità, ha degli attanti. I quasi-predicati denotano entità che per natura (o per funzione) sono legate ad altre entità, poiché il loro stesso senso non può essere inteso se non in relazione ad altre entità (Mel'čuk, 2012). Sono esempi tipici di quasi-predicati: i nomi di professioni, come *professore* (=persona x che insegna la materia y a z); i sostantivi che denotano relazioni personali, familiari, come *sorella, padre, fidanzato* (perché implicano una relazione con qualcuno: una sorella è sorella *di qualcuno*, essere sorella non è una qualità intrinseca di una persona) o anche animali domestici (=animali identificati primariamente per il fatto di essere animali *di qualcuno*); parti del corpo; parti di un tutto; sostantivi che denotano artefatti, cioè oggetti creati dall'uomo e spesso dotati di una funzione specifica, come *coltello* (=oggetto che serve affinché una persona x tagli y), nomi di istituzioni, nomi collettivi come *folla, flotta, gregge*, etc. (Cfr. cap.1 sez. 1.5.2.1.1). Per approfondimenti si rimanda a Polguère (2012).

rappresentazione  
linguistica

Rappresentazione di un enunciato a un determinato livello di analisi linguistica (ad esempio la rappresentazione semantica di un enunciato). Nella teoria Senso-Testo una rappresentazione linguistica è definita come un insieme di oggetti formali chiamati *strutture*. Tra queste è

---

<sup>45</sup> neologismo da me coniato per lo scopo di questo lavoro come traduzione del termine inglese *wordform*, proprio della teoria Senso-Testo. La letteratura Senso-Testo in lingua francese utilizza il termine *mot-forme*.

possibile distinguere una struttura centrale, che riflette le caratteristiche salienti dell'enunciato per il livello considerato, e una serie di strutture periferiche, portatrici di altre informazioni (Mel'čuk, 2012) (cfr. cap. 1 sez. 1.5.2).

Ad esempio, la rappresentazione semantica è costituita da una struttura semantica (che costituisce la struttura centrale) più una serie di altre strutture periferiche: la struttura comunicativa (relativa alla distinzione tema/rema), la struttura retorica, e la struttura referenziale (cfr. cap. 1 sez. 1.5.2.1).

relazione lessicale

Relazione esistente tra due lessemi di una data lingua, basata (solitamente, ma non necessariamente) su una relazione semantica o sintattica (Wanner, 1996, Mel'čuk, 1996).

Esempi di relazioni lessicali sono la sinonimia, l'antonimia, la conversività (ossia la relazione tra due conversivi come *padre* e *figlio*), la derivazione semantica, la derivazione sintattica, la relazione di co-occorrenza condizionata (nel caso delle collocazioni) (cfr. cap.2 sez. 2.3.2).

segno linguistico

La teoria Senso-Testo identifica il segno linguistico come un segno avente una tripla articolazione: significato + significante + proprietà combinatorie (cfr. cap.1 sez. 1.4).

Oltre al significato e al significante (intesi nell'accezione di Saussure), Mel'čuk (2012) individua un'ulteriore articolazione del segno linguistico, quella costituita dalle **proprietà combinatorie** del segno, ossia i diversi modi in cui esso può combinarsi ad altri segni linguistici all'interno di una data lingua naturale. In termini tecnici, le proprietà combinatorie di un segno sono l'insieme delle restrizioni (più o meno arbitrarie) sulle co-occorrenze del segno linguistico (cfr. cap. 1 sez. 1.4 e cap. 2 sez. 2.3.2).

All'interno del Dizionario esplicativo-combinatorio (cfr. cap. 1 sez. 1.6.1), la parte dedicata alle proprietà combinatorie di un segno linguistico include:

- la corrispondenza tra attanti semantici e attanti sintattici profondi;
- sinonimi, antonimi e conversivi;
- co-occorrenze ristrette (collocazioni), espresse in termini di Funzioni Lessicali.

semantema

Detto anche unità semantica, un semantema è l'unità di base della rappresentazione semantica (Mel'čuk, 2012). I semantemi corrispondono ai nodi dei grafi che costituiscono la struttura semantica. "Un semantema è il senso di un'unità lessicale completa – cioè di un lessema o di una locuzione di una lingua" (Mel'čuk, 2012:2015). (Cfr. cap.1 sez. 1.5.2.1.1).

senso	<p>Contenuto mentale che il mittente esprime attraverso segnali linguistici fisici, che possono essere ricevuti e compresi dal destinatario.</p> <p>Nella teoria Senso-Testo il senso è inteso unicamente come <i>senso linguistico</i>, ossia il contenuto informativo che può essere ricavato da un'espressione esclusivamente mediante conoscenze linguistiche, escludendo tutti gli aspetti extralinguistici (pragmatici, culturali, etc.) (Mel'čuk, 2012).</p> <p>La teoria Senso-Testo non si occupa di definire che cosa sia il senso in sé (poiché il senso in sé è una nozione non direttamente accessibile), ma dà una definizione di senso derivata dal concetto di "stesso senso": <b>il senso linguistico di un'espressione è la proprietà comune a tutte le parafrasi (o sinonimi) di tale espressione.</b> (Cfr. cap. 1 sez. 1.4.1 e sez. 1.5.2.1).</p> <p>Allo stesso modo, la teoria Senso-testo non si occupa né degli aspetti legati al rapporto tra senso e referente, né degli aspetti psicolinguistici o legati alla genesi del senso a livello neurologico.</p> <p>La teoria Senso-Testo si serve di una rappresentazione simbolica del senso, detta rappresentazione semantica (RSem), che costituisce uno dei pilastri del modello Senso-Testo. (Cfr. cap. 1 sez. 1.5.2.1).</p>
testo	<p>Segnale linguistico fisico – grafico, acustico o gestuale –utilizzato dal mittente per esprimere un senso, e che può essere ricevuto e compreso da un destinatario (Mel'čuk, 2012). In altre parole nella teoria Senso-Testo il testo è inteso semplicemente come significante linguistico, e può riferirsi a segmenti di varia grandezza: morfemi, parole, sintagmi, frasi o addirittura interi paragrafi (Cfr. cap. 1 sez. 1.4.1).</p> <p>La teoria Senso-Testo prende in considerazione una rappresentazione simbolica dei testi, che convenzionalmente coincide con la rappresentazione fonologica (Mel'čuk, 2013:113).</p>
unità lessicale	<p>Categoria lessicale che comprende i lessemi e le locuzioni (Mel'čuk, 2012). Lessemi e locuzioni sono unità lessicali nel senso che costituiscono le unità di base del lessico, in quanto espressioni dotate di un significato lessicale.</p> <p>Oltre a essere una categoria lessicale e lessicografica dell'approccio Senso-Testo, le unità lessicali compaiono anche nel modello Senso-Testo a livello sintattico profondo: lessemi e locuzioni costituiscono infatti i nodi della struttura sintattica profonda e in questo contesto sono dette quindi <i>unità lessicali profonde</i> (cfr. cap. 1 sez. 1.5.2.2).</p>

vocabolo

Termine proprio della lessicografia esplicativa-combinatoria (Mel'čuk e al., 1999); categoria lessicale che designa un insieme di unità lessicali che:

- hanno la stessa forma
- hanno proprietà combinatorie simili
- hanno un legame semantico importante

La nozione di vocabolo è legata alla nozione di *polisemia*. Tradizionalmente, una parola polisemica è una parola che possiede diversi significati legati tra loro. Per la linguistica Senso-Testo, la categoria "parola polisemica" è rappresentata dalla categoria "vocabolo", mentre le singole accezioni corrispondono alla categoria "unità lessicale" (ricordiamo che un'unità lessicale è un'espressione dotata di un singolo significato).

Ad esempio *radio* (= apparecchio di ricezione radiofonica) e *radio* (= emittente radiofonica) costituiscono due unità lessicali diverse, ma fanno parte dello stesso vocabolo.

Quanto appena affermato non vale - ovviamente - per i casi di omonimia, poiché due parole omonime non sono legate da un nesso semantico, e pertanto costituiscono due vocaboli indipendenti. Ad esempio *radio* (= apparecchio di ricezione radiofonica) e *radio* (= osso del braccio) costituiscono due vocaboli diversi.



## Appendice 2: Funzioni Lessicali Standard Semplici

Le funzioni lessicali standard semplici (simple standard lexical functions) sono qui raccolte in una tabella che costituisce un'esposizione schematica del contenuto di Mel'čuk (1996), parzialmente rielaborato dalla sottoscritta nei seguenti aspetti:

- la terminologia specifica è stata adattata all'italiano (Cfr. cap. 1, sez. 1.2);
- La maggior parte degli esempi sono stati ideati dalla sottoscritta per la lingua italiana. Sono stati riportati anche esempi in lingua francese e inglese, tratti da Mel'čuk (1996), perché ritenuti particolarmente efficaci. Gli esempi in lingua inglese o francese sono preceduti dalla sigla EN o FR;
- le spiegazioni originarie di Mel'čuk – che spesso presuppongono la conoscenza del modello Senso-Testo e le sue convenzioni teoriche – sono state rielaborate per renderle più sintetiche e comprensibili anche a chi non ha familiarità col modello Senso-Testo;
- si è scelto di non avvalersi della distinzione tra funzioni lessicali sintagmatiche e paradigmatiche, poiché per alcune funzioni lessicali tale distinzione è poco netta;
- sono state adottate le seguenti convenzioni grafiche:
  - nella tabella il termine *funzione lessicale* è abbreviato in FL
  - i nomi delle funzioni lessicali sono riportati nella seconda colonna, in carattere minuscolo e con iniziale maiuscola;
  - negli esempi, la base è riportata tra parentesi tonde; le parentesi graffe sono utilizzate invece per racchiudere l'insieme dei possibili valori di una funzione lessicale. Esempio: S1(insegnare) = {insegnante, maestro, professore, ...};
  - nei casi in cui il valore della funzione lessicale è un collocato della base, nello spazio riservato al valore della funzione lessicale comparirà anche, tra parentesi quadre, il simbolo ~ (che rappresenta la base) e l'articolo, la preposizione o l'eventuale altro determinante (indicato come DET) che lega base e collocato. Esempio: S<sub>instr</sub>(delitto) = arma [del ~];
  - nei casi in cui negli esempi siano presenti gli attanti della base, essi sono indicati mediante la seguente notazione: X=primo attante; Y=secondo attante; Z=terzo attante. Il simbolo N indica invece un elemento nominale che co-occorre col valore della funzione;
  - nei casi in cui una funzione lessicale abbia come valori sia collocati della base sia parole che sostituiscono la base in contesto di frase (concetto di  *fusione*, cfr. cap. 2 sez. 2.4.2.2), queste ultime sono riportate dopo una doppia barra obliqua. Esempio: IncepOper1(malattia) = contrarre [una ~] // ammalarsi

Alcune funzioni lessicali necessiterebbero di una trattazione più approfondita – ad esempio le funzioni lessicali verbali, per le quali occorre specificare la struttura degli attanti sintattici. Per l'approfondimento di questi aspetti si rimanda a (Mel'čuk, 1996).

n.	FL		spiegazione	esempi
1	Syn / QSyn	Sinonimo	<p>Il valore della funzione presenta le seguenti caratteristiche in comune con la base:</p> <ul style="list-style-type: none"> <li>-Stesso senso</li> <li>-Stessa parte del discorso</li> <li>-Stessa struttura di attanti sintattici<sup>46</sup></li> </ul> <p>Sono possibili variazioni stilistiche o di registro.</p> <p>Dal momento che la sinonimia perfetta è rara nelle lingue, si tende a preferire l'impiego della FL QSyn, che indica un quasi-sinonimo.</p>	Syn(sedia) = seggiola
		Quasi-sinonimo	<p>Il valore della funzione ha un senso leggermente differente, rappresentato per mezzo di simboli</p> <p>QSyn<math>\supset</math> = senso più specifico  QSyn<math>\subset</math> = senso più generico  QSyn<math>\cap</math> = intersezione di senso, un senso che ha solo alcune componenti in comune con il senso della base.</p>	<p>Syn<math>\supset</math>(firmare) = autografare  Syn<math>\subset</math>(assalto) = attacco  Syn<math>\cap</math>(arrestare) = fermare</p>
2	Anti	Antonimo	<p>Stessa parte del discorso della base, ma il senso è la negazione, il contrario del senso della base.</p> <p>Esistono diversi tipi di antonimia, che qui non citiamo, per la cui trattazione rimandiamo a Gagner e L'Homme (2016).</p>	Anti(chiuso) = aperto
3	Conv <sub>kij</sub>	Conversivo	<p>Stessa parte del discorso della base, ma il senso rappresenta la "controparte" del senso della base. Gli indici numerici descrivono l'ordine degli attanti. Perché la funzione Conv sia applicabile, la base deve</p>	<p>Conv<sub>21</sub>(temere) = fare paura  Conv<sub>21</sub>(moglie) = marito  Conv<sub>321</sub>(vendere) = comprare</p>

<sup>46</sup> Cfr. cap. 1 sez. 1.5.2.1.1 e la voce **attante**<sup>α</sup> nel glossario in Appendice 1.

			avere almeno due attanti semantici.	
4	Contr	Contrastivo	Parola che nell'uso è associata alla base come suo opposto, senza tuttavia rappresentare un antonimo perché non comporta una negazione della base.	Contr(anima) = corpo Contr(cielo) = terra Contr(forma) = sostanza
5	Epit	Epiteto	Il valore della funzione è solitamente un aggettivo che nell'uso accompagna la base riprendendone una componente semantica, senza aggiungere sensi ulteriori.	EN: Epit(winner) = lucky [~] (speaking of lottery) EN: Epit(oaf) = clumsy [~] <sup>47</sup>
6	Gener	Iperonimo	Iperonimo della base, nello specifico "the closest generic concept" (Mel'čuk, 1996:51).	Gener(gatto) = felino
7	Figur	Metafora	Il valore della funzione è una metafora con cui nell'uso ci si riferisce alla base, e che spesso accompagna la base stessa.	Figur(passione) = fuoco [della ~] EN: Figur(fog) = wall [of ~]
8	S <sub>0</sub>	Derivato sintattico nominale	Derivati sintattici della base. <sup>48</sup>	S <sub>0</sub> (accusare) = accusa S <sub>0</sub> (buono) = bontà S <sub>0</sub> (tardi) = ritardo
9	A <sub>0</sub>	Derivato sintattico aggettivale		A <sub>0</sub> (città) = {urbano, cittadino} A <sub>0</sub> (includere) = inclusivo
10	V <sub>0</sub>	Derivato sintattico verbale		V <sub>0</sub> (analisi) = analizzare V <sub>0</sub> (migliore) = migliorare
11	Adv <sub>0</sub>	Derivato sintattico avverbiale		Adv <sub>0</sub> (gentile) = gentilmente
12	S <sub>i</sub>	Attante (Derivato semantico nominale)	Il valore della funzione è un derivato semantico della base, cioè un sostantivo <b>standard</b> che esprime l'attante semantico <i>i</i> della base. Ad esempio: omicidio ha 2 attanti semantici: (1) la	S <sub>2</sub> (omicidio) = vittima S <sub>1</sub> (omicidio) = {omicida, assassino, autore [dell'~], ...}  S <sub>1</sub> (insegnare) = {insegnante, maestro, professore, ...}

<sup>47</sup> Esempi tratti da Mel'čuk (1996).

<sup>48</sup> Per le funzioni S<sub>0</sub>, A<sub>0</sub>, V<sub>0</sub>, Adv<sub>0</sub> la base della funzione deve necessariamente appartenere a una parte del discorso diversa da quella espressa nel nome della funzione.

			<p>persona che uccide; (2) la persona che viene uccisa.</p> <p>Insegnare ha 3 attanti: (1) la persona che insegna, (2) ciò che viene insegnato, (3) chi riceve l'insegnamento.</p> <p>È importante comprendere la differenza rispetto ai derivati sintattici: alcuni derivati semantici possono coincidere con derivati sintattici (<math>S_1</math>(omicidio) = omicida), ma la differenza appare chiara nell'esempio <math>S_1</math>(omicidio) = assassino, in cui non c'è legame derivazionale.</p>	$S_3$ (insegnare) = {studente, allievo, alunno, ...}
13	$S_{instr}$	Strumento (Derivato semantico relativo a un circostante) <sup>49</sup>	Sostantivo standard che esprime lo strumento con cui si realizza la situazione espressa dalla base.	$S_{instr}$ (delitto) = arma [del ~]
14	$S_{med}$	Mezzo (Derivato semantico relativo a un circostante)	Sostantivo standard che esprime il mezzo con cui si realizza la situazione espressa dalla base.	$S_{med}$ (scuse) = lettera [di ~] <sup>50</sup>
15	$S_{mod}$	Modo (Derivato semantico relativo a un circostante)	Sostantivo standard che esprime il modo, il carattere di attuazione della situazione espressa dalla base.	$S_{mod}$ (vita) = stile [di ~]
16	$S_{loc}$	Luogo (Derivato semantico relativo a un circostante)	Sostantivo standard che esprime il luogo in cui si realizza la situazione espressa dalla base.	$S_{loc}$ (battaglia) = campo [di ~] $S_{loc}$ (corsa) = pista [da ~]
17	$S_{res}$	Risultato (Derivato semantico relativo a un circostante)	Sostantivo standard che esprime il risultato della situazione o azione espressa dalla base.	$S_{res}$ (riferire) = rapporto
18	$Able_i$		Il valore della funzione è un aggettivo che esprime l'attributo tipico dell'attante $i$	EN: $Able_1$ (to cry) = tearful

<sup>49</sup> I circostanti, a differenza degli attanti, sono elementi facoltativi della situazione descritta dalla base.

<sup>50</sup> Esempio tratto dal DiCouèbe (<http://olst.ling.umontreal.ca/dicouebe/main.php>).

			della base <b>rispetto al suo ruolo potenziale nella situazione</b> rappresentata dalla base.	
19	Qual <sub>i</sub>		<p>Aggettivo che esprime una qualità <b>intrinseca e standard</b> (nell'uso) dell'attante <i>i</i> della base, indipendentemente dal suo ruolo nella situazione.</p> <p>Qual<sub>i</sub> non implica necessariamente Able<sub>i</sub>, ma solo la sua probabilità.</p>	EN: Qual <sub>1</sub> (to cry) = sad
20	Sing		Il valore della funzione è un sostantivo con cui ci si riferisce a un'istanza o un'unità di un insieme o di un elemento composto di tante unità (rappresentato dalla base della funzione).	<p>Sing(flotta) = nave</p> <p>Sing(sabbia) = granello</p>
21	Mult	Nome collettivo	Sostantivo con cui ci si riferisce a un insieme formato da istanze della base.	<p>EN: Mult(sheep) = herd</p> <p>Mult(soldato) = esercito</p>
22	Cap		Sostantivo con cui ci si riferisce all'autorità a capo dell'elemento espresso dalla base.	<p>Cap(università) = rettore</p> <p>Cap(scuola) = preside</p> <p>Cap(orchestra) = direttore</p>
23	Equip		Sostantivo con cui ci si riferisce al personale specifico che lavora nel contesto indicato dalla base	<p>Equip(teatro) = compagnia</p> <p>Equip(nave) = equipaggio</p> <p>Equip(film) = {cast, troupe...}</p>
24	A <sub>i</sub>	Derivato semantico aggettivale	<p>Il valore della funzione esprime l'aggettivo standard che descrive la qualità dell'attante sintattico profondo <i>i</i> della base, nella situazione descritta dalla base.</p> <p>Ad esempio, il sintagma verbale <i>fare paura</i> ha due attanti: (1) chi o cosa fa paura (2) chi o cosa prova paura di conseguenza.</p> <p>Anche in questo caso, ci sono evidenti sovrapposizioni con i derivati sintattici aggettivali, ma la derivazione semantica è di natura diversa, e opera una distinzione più fine.</p>	<p>A<sub>1</sub>(fare paura) = pauroso</p> <p>A<sub>2</sub>(fare paura) = impaurito</p> <p>A<sub>2</sub>(conoscere) = familiare</p>

25	Adv <sub>i</sub>	Derivato semantico avverbiale	Il valore della funzione è un avverbio che descrive l'azione o lo stato dell'attante <i>i</i> della base, in relazione al suo ruolo nella situazione definita dalla base.	Adv <sub>1</sub> (rabbia) = con [~] // rabbiosamente Adv <sub>1</sub> (arrendersi) = nell'atto di [~] Adv <sub>2</sub> (attaccare) = sotto attacco
26	Imper	Ordine	Il valore della funzione è il comando standard con cui si ordina di compiere l'azione rappresentata dalla base.	Imper(sparare) = fuoco! Imper(entrare) = avanti! Imper(girare[riprese]) = azione! EN: Imper(to seize arms) = take up arms!
27	Result	Risultato	Il valore della funzione è un verbo che esprime il risultato dell'azione/processo corrispondente alla base.	Result(comprare) = possedere

Le Funzioni lessicali che seguono possono essere distinte in FL nominali, aggettivali, avverbiali, verbali o preposizionali a seconda della parte del discorso del valore di ciascuna (Mel'cuk, 1996).

28	Centr	Centro	Il valore della funzione è la parola standard con cui ci si riferisce al centro (anche metaforico) della base.	Centr(forest) = the thick [of the ~] Centr(desert) = the heart [of the ~] Centr(life) = prime [of one's ~]
29	Magn	Intensità	Il valore della funzione è un'espressione aggettivale o avverbiale che esprime intensificazione del senso della base. Magn si trova spesso in funzioni lessicali complesse.	Magn(pioggia) = battente Magn(bagnato) = {[~] fradicio, [~] come un pulcino, ...} Magn(silenzio) = {[~] assoluto, [~] di tomba, ...} Magn(gridare) = [~] a squarciagola AntiMagn(maggioranza) = [~] risicata
30	Plus	Più	La funzione Plus compare quasi esclusivamente all'interno di funzioni lessicali complesse, in cui apporta il senso 'aumento'.	PredPlus(prezzi) = {aumentare, salire, ...}
31	Minus	Meno	Compare quasi esclusivamente all'interno di funzioni lessicali complesse, in cui apporta il senso 'diminuzione'.	CausPredMinus(colpo) = attutire [ART ~]

32	Ver	Ideale	Il senso apportato alla base dal valore della funzione è 'così come dovrebbe essere'.	Ver(sentenza) = [~] giusta Ver(punizione) = [~] meritata Ver(salute) = buona [~]
33	Bon	Buono	Il valore della funzione è un'espressione aggettivale o avverbiale che significa 'buono' o 'bene', in genere rappresenta un collocato più forte e più specifico nella lingua L rispetto al semplice 'buono'. Si trova spesso in funzioni lessicali complesse in combinazione con Anti, Pred, Plus, ecc.	Bon(proposta) = [~] allettante Bon(alunno) = {[~] diligente, [~] brillante, [~] dotato...} AntiBon(odore) = {cattivo [~], [~] nauseabondo...}
34	Pos <sub>i</sub>	Positivo	Il valore della funzione esprime il senso 'positivo', 'di segno positivo', in genere non riferito alla base, ma all'operato dell'attante <i>i</i> da parte dell'attante 1. Per chiarire la differenza rispetto a Bon, si ricorda che il senso di Bon agisce sulla base (senza riferirsi alla struttura di attanti). Proponiamo un esempio contrastivo con la base <i>recensione</i> , i cui attanti sono (1) l'autore e (2) la cosa recensita. Anche Pos compare spesso all'interno di funzioni lessicali complesse.	Pos <sub>2</sub> (recensione) = {[~] entusiasta, lusinghiera} vs. Bon(recensione) = {[~] ben scritta, [~] acuta, [~] interessante, ...}
35	Loc <sub>in</sub>	Ubicazione spaziale (in)	Funzione lessicale preposizionale il cui valore è una preposizione di stato nel luogo indicato dalla base. La preposizione governa la base.	EN: Loc <sub>in</sub> (height) = at [a ~ of ...] Loc <sub>in</sub> (teatro) = a [~] Loc <sub>in</sub> (montagna) = in [~] Loc <sub>in</sub> (mare) = al [~] Loc <sub>in</sub> (scuola) = a [~] Loc <sub>in</sub> (ufficio) = in [~]
36	Loc <sub>ab</sub>	Ubicazione spaziale (da)	Preposizione di moto da luogo o che esprime una prospettiva spaziale. Da un punto di vista sintattico, la base dipende dalla preposizione.	EN: Loc <sub>ab</sub> (height) = from [a ~ of ...]

37	Loc <sub>ad</sub>	Ubicazione spaziale (verso)	Preposizione di moto verso luogo. Da un punto di vista sintattico, la base dipende dalla preposizione.	EN: Loc <sub>ad</sub> (height) = to [a ~ of ...]
38	Loc <sup>temp</sup> <sub>in</sub>	Contesto temporale (in)	Preposizione (o espressione complessa dal valore preposizionale) temporale specifica per indicare il contesto temporale.	Loc <sup>temp</sup> <sub>in</sub> (primavera) = in [~] FR: Loc <sup>temp</sup> <sub>in</sub> (printemps) = au [~] Loc <sup>temp</sup> <sub>in</sub> (Natale) = a [~]  EN: Loc <sup>temp</sup> <sub>in</sub> (arrest) = at the time of [DET ~] EN: Loc <sup>temp</sup> <sub>in</sub> (game) = during [DET ~] EN: Loc <sup>temp</sup> <sub>in</sub> (communism) = under [~]
39	Instr	Strumento	Il valore della funzione è la preposizione (o l'espressione preposizionale) che regge la base.	EN: Instr(mail) = by [~] EN: Instr(satellite) = via [~] EN: Instr(typewriter) = on [DET ~]
40	Propt	Causa (dal latino <i>propter</i> , a causa di)	Preposizione (o espressione preposizionale) di causa.	EN: Propt(fear) = out of [~]
41	Copul	Copula	Il valore della funzione è la <i>copula</i> che unisce il soggetto col suo predicato. In genere nelle lingue occidentali la copula è un verbo, in italiano tipicamente è il verbo <i>essere</i> .	Copul(esempio) = essere [un ~]
42	Pred	Predicato	Il valore della funzione è un'espressione fusa formata dalla copula e della base della funzione. Spesso <i>Pred</i> si trova combinata con altre funzioni lessicali per formare funzioni lessicali complesse. Anche in questo caso notiamo alcune coincidenze tra valori della funzione <i>Pred</i> e valori della funzione <i>V<sub>o</sub></i> .	Pred(esempio) = esemplificare Pred(insieme) = accompagnare EN: Pred(close) = to neighbor IncepPred(malato) = ammalarsi
43	Oper <sub>i</sub>	Verbo di supporto il cui attante è la base	Il valore della funzione è un <b>verbo di supporto</b> (ossia "vuoto" semanticamente, la cui sola funzione è quella di "reggere" la base) che ha le seguenti caratteristiche:	<b>Oper<sub>1</sub></b> (consiglio) = dare [un ~] <b>Oper<sub>2</sub></b> (consiglio) = ricevere [un ~]

			<p>- Il suo primo attante sintattico profondo (cioè il soggetto in sintassi di superficie) è l'attante semantico<sup>51</sup> <i>i</i> della base</p> <p>- il <b>secondo attante</b> sintattico profondo (il complemento) <b>è la base stessa</b> (cfr. cap. 2 sez. 2.5.1.1)</p> <p>Si trova spesso associato ad altre funzioni lessicali nella formazione di funzioni lessicali complesse e configurazioni di funzioni lessicali.</p>	<p>[consiglio ha due attanti sintattici profondi: (X) chi dà il consiglio, (Y) chi lo riceve)]</p>
44	Func <sub>0/i</sub>	Verbo di supporto il cui I attante è la base	<p>Il valore della funzione è un <b>verbo di supporto</b> di cui:</p> <p>- il <b>primo attante</b> sintattico profondo <b>è la base</b>,</p> <p>- l'eventuale secondo attante sintattico profondo corrisponde all'attante semantico <i>i</i> della base (cfr. cap. 2 sez. 2.5.1.2).</p> <p>Anche Func si trova spesso associato ad altre funzioni lessicali nella formazione di funzioni lessicali complesse e configurazioni di funzioni lessicali.</p>	<p><b>Func<sub>0</sub>(vento) = [~] soffiare</b> <i>Il vento soffia</i></p> <p>Si ha Func<sub>0</sub>, quando la base non ha altri attanti (come nell'esempio <i>il vento soffia: vento</i> non ha attanti) oppure quando gli altri attanti della base non diventano attanti del verbo:</p> <p><b>Func<sub>0</sub>(offerta) = [~] valere</b> <i>L'offerta vale fino al 15 maggio.</i></p> <p>La parola <i>offerta</i> ha 3 attanti semantici, come il verbo offrire: (1) chi offre, (2) la cosa offerta, (3) la persona a cui si offre la cosa. Nella frase sopra, tuttavia, <u>questi attanti non fanno parte della struttura attanziale del verbo <i>valere</i></u>. Il verbo resta pertanto rappresentato dalla funzione Func<sub>0</sub>.</p> <p><b>Func<sub>1</sub>(decisione) = [~] arrivare</b> <i>La decisione arriva dalla dirigenza.</i></p>

<sup>51</sup> in realtà nella definizione classica delle funzioni lessicali data da Mel'čuk (1996), gli indici numerici delle funzioni lessicali Oper, Func, Labor, Real, Fact e Labreal non si riferiscono agli attanti semantici della base, bensì agli attanti sintattici profondi della base definiti nella diatesi di base della parola (Mel'čuk, 1996). Nella presente esposizione, in accordo con il professor Lareau, si è deciso di optare per una definizione leggermente diversa, che prende in considerazione gli attanti semantici (cfr. cap. 2 sez. 2.5).

				<p>il primo attante di <i>decisione</i> è chi decide, in questo caso la dirigenza.</p> <p><b>Func<sub>2</sub></b>(punizione) = [~] colpire  <i>La punizione divina colpì tutto il popolo egiziano.</i>  Punizione ha 2 attanti semantici, come il verbo punire: (1) chi punisce e (2) chi viene punito.</p> <p><b>Func<sub>3</sub></b>(discussione) = [~] riguardare  <i>La discussione di oggi riguarda il bilancio.</i>  Discussione ha 3 attanti semantici, come il verbo discutere: (1) chi discute, (2) la persona con cui discute, (3) la cosa di cui si discute.</p>
45	Labor <sup>r<sub>ij(k)</sub></sup>	Verbo di supporto il cui III attante è la base	<p>Il valore della funzione è un <b>verbo di supporto</b> con la seguente struttura attanziale:</p> <ul style="list-style-type: none"> <li>- il suo primo attante sintattico profondo (o soggetto) è l'attante semantico <i>i</i> della base</li> <li>- il secondo attante (o complemento) è l'attante semantico <i>j</i> della base</li> <li>- <b>il terzo attante è la base</b></li> <li>- il quarto attante è l'eventuale attante semantico <i>k</i> della base (cfr. cap. 2 sez. 2.5.1.3).</li> </ul>	<p><b>Labor<sub>12</sub></b>(interrogatorio)= sottoporre [Y a DET ~]  <i>La polizia sottoporrà Battisti a un interrogatorio.</i>  Interrogatorio ha 2 attanti semantici, come il verbo <i>interrogare</i>: (1) chi interroga, (2) chi viene interrogato.</p> <p><b>Labor<sub>12</sub></b>(eredità) = ricevere [Y in ~]  <i>I nipoti hanno ricevuto una casa in eredità.</i></p> <p><b>Labor<sub>321</sub></b>(eredità) = lasciare [Y in ~ a Z]  <i>Il vecchio lasciò una casa in eredità ai suoi nipoti.</i>  Eredità ha 3 attanti semantici, come il verbo ereditare: (1) chi eredita, (2) la cosa ereditata, (3) la persona da cui si eredita.</p>
46	Incep	Inizio Verbo di fase	Fa parte del gruppo delle <b>Phasal Lexical Functions</b> , funzioni lessicali che	<p><b>Incep</b>(dormire) = addormentarsi  <b>IncepOper<sub>1</sub></b>(malattia) = contrarre [una ~] // ammalarsi</p>

			<p>esprimono fasi di un'azione o di uno stato.</p> <p>La base di una Phasal Lexical function può essere un verbo o un'espressione verbale, a cui il valore della funzione va a sostituirsi nel testo.</p> <p>In caso si voglia applicare la una Phasal Lexical Function a una base nominale, sarà pertanto necessario l'intervento di un verbo di supporto (<i>Func</i>, <i>Oper</i> o <i>Labor</i>) o di realizzazione (<i>Fact</i>, <i>Real</i> o <i>Labreal</i>), dando così luogo a una funzione lessicale complessa come nell'esempio <i>IncepOper<sub>1</sub>(love)</i>.</p> <p>Il valore della funzione <i>Incep</i> è un verbo il cui senso è "iniziare (a) [senso dalla base]".</p>	<p><b>IncepFunc<sub>0</sub></b>(sole) = [~] sorgere</p> <p><b>IncepFunc<sub>0</sub></b>(guerra) = [~] scoppiare</p>
47	Cont	<p>Continuazione</p> <p>Verbo di fase</p>	<p><b>Phasal lexical function</b> il cui valore è un verbo che esprime il senso "continuare (a) [senso della base]".</p> <p>Valgono le specificità e le regole delle Phasal lexical functions: possono essere applicate solo a una base verbale o a una funzione lessicale verbale.</p>	<p>EN: <i>ContFunc<sub>0</sub></i>(offer) = [~] stands</p> <p><i>ContOper<sub>1</sub></i>(calma) = mantenere [la ~]</p>
48	Fin	<p>Fine</p> <p>Verbo di fase</p>	<p><b>Phasal lexical function</b> il cui valore è un verbo che esprime il senso "cessare (di) [senso della base]".</p>	<p><i>Fin</i>(ricordare) = dimenticare</p> <p>FR: <i>FinFunc<sub>0</sub></i>(craintes) = se dissiper</p> <p>EN: <i>FinOper<sub>0</sub></i>(power) = to lose [one's ~ over N]</p>
49	Caus <sub>i</sub>	<p>Causa dell'inizio</p>	<p>Funzione lessicale causativa. Il valore della funzione è un verbo che esprime una causa, più precisamente "do something so that a situation <b>begins</b> occurring" (Mel'čuk, 1996:65).</p> <p>Come le altre funzioni lessicali causative (Perm e</p>	<p>FR: <b>CausFunc<sub>0</sub></b>(faim) = donner [~]</p> <p><i>La fatigue donne faim</i></p> <p>= la fatica fa venire fame</p> <p>= la fatica causa il fatto che la fame <u>si manifesti</u> (Func<sub>0</sub>)</p> <p>FR: <b>CausOper<sub>1</sub></b>(meurtre) = pousser [quelqu'un au ~]</p>

			<p>Liqu), <b>si trova perlopiù in FL complesse</b> associata a FL verbali, anche se in alcune lingue (tra cui il russo) <i>Caus</i> può essere applicata autonomamente a una base. Come ogni FL causativa (cfr. Mel'čuk, 1996), <i>Caus</i> modifica la struttura attanziale della base aggiungendo un attante "esterno" (cioè l'agente della causa), tuttavia esistono casi in cui l'agente della causa coincide con uno degli attanti semantici della base: in tal caso, esso viene indicato mediante un indice numerico nel nome della funzione.</p>	<p><i>La jalousie a poussé Paul au meurtre</i>  = [lett.] la gelosia ha spinto Paul all'omicidio  = la gelosia ha spinto Paul a <i>commettere</i> (Oper<sub>1</sub>) un omicidio.</p> <p><b>Caus</b>(tela)= tessere [DET ~]<sup>52</sup></p> <p><b>Caus<sub>2</sub>Func<sub>2</sub></b>(attenzione) = attirare [l'~]  <i>Il danzatore ha attirato l'attenzione dei critici.</i></p> <p>Attenzione ha due attanti: (1) chi presta attenzione, (2) <u>la cosa a cui si presta attenzione</u>.</p> <p>Osserviamo ora la seguente parafrasi della frase precedente:  <i>Il danzatore ha causato il fatto che l'attenzione dei critici fosse</i> [func<sub>2</sub>] su di <i>lui</i>.</p> <p>L'attante 2 della base è quindi anche l'agente della causa dell'azione legata alla base: ecco perché abbiamo Caus<sub>2</sub>.</p>
50	Perm <sub>i</sub>	Causa della continuazione	<p>Funzione lessicale causativa. Il valore della funzione è un verbo che esprime la causa della continuazione della situazione espressa dalla base, o più precisamente il fatto di <b>non causare la fine della situazione: "do nothing which would cause that a situation stops occurring"</b> (Mel'čuk, 1996:65).</p> <p>Anche per <i>Perm</i> valgono le restrizioni e le particolarità delle FL causative.</p>	<p><b>Perm<sub>2</sub>Func<sub>2</sub></b>(offese) = sopportare [DET ~]  <i>L'uomo saggio sopporta pazientemente le offese.</i></p> <p><i>Offesa</i> ha due attanti: (1) chi offende, (2) <u>chi è offeso</u>.</p> <p>Osserviamo ora la seguente parafrasi della frase precedente:  <i>L'uomo saggio NON causa la fine delle offese.</i></p> <p>L'attante 2 della base è quindi anche l'agente della causa dell'azione legata alla base: ecco perché abbiamo Caus<sub>2</sub>.</p>

<sup>52</sup> Esempio tratto da Gatti (1992:190)

51	Liqu <sub>i</sub>	Causa della fine	<p>Funzione lessicale causativa. Il valore della funzione è un verbo che esprime la <b>causa della fine</b> della situazione espressa dalla base, o più precisamente <b>“do something so that a situation stops occurring”</b> (Mel’čuk, 1996:65).</p> <p>Anche per <i>Liqu</i> valgono le restrizioni e le particolarità delle FL causative.</p>	<p>Liqu(dormire) = svegliare [qualcuno]  LiquFunc<sub>0</sub>(tracce) = cancellare [le ~]  LiquFact<sub>0</sub>(legge) = abrogare [DET ~]</p>
52	Real <sub>i</sub>	Verbo di realizzazione il cui II attante è la base	<p>Il valore della funzione è un verbo che esprime la <b>realizzazione della funzione</b> propria della base.</p> <p>Ciò implica che la base di <i>Real</i> (così come di <i>Fact</i> e <i>Labreal</i>) sia un’entità che possiede una funzione precisa, o un’implicita idea di realizzazione.</p> <p>Per quanto riguarda la struttura attanziale e l’indice numerico, <i>Real</i> funziona come <i>Oper</i> (possiamo dire che ne sia l’equivalente, ma con il senso di realizzazione):</p> <ul style="list-style-type: none"> <li>- Il suo primo attante sintattico profondo (cioè il soggetto in sintassi di superficie) è l’attante semantico <i>i</i> della base</li> <li>- il <b>secondo attante</b> sintattico profondo (il complemento) <b>è la base stessa.</b></li> </ul> <p>Come <i>Oper</i>, e come le altre FL verbali, si trova spesso associato ad altre FL nella formazione di FL complesse e configurazioni di FL.</p>	<p><b>Real<sub>1</sub></b>(camicia) = indossare [DET ~]  <i>Federico indossa una camicia</i></p> <p><i>Camicia</i> è una parola che designa un <b>artefatto</b> che possiede una funzione, e nel modello senso-testo è definita un <b>quasi-predicato</b>. La sua struttura di attanti, non del tutto intuitiva, è desunta dalla definizione che ne dà il Dizionario Esplicativo-Combinatorio: “oggetto che serve affinché X [=attante 1] la indossi”. L’attante 1 di camicia è dunque la persona che la indossa. Ecco perché <b>Real<sub>1</sub></b>. Inoltre, la definizione esplicita l’esistenza di una funzione precisa.</p> <p>Dunque, tornando all’esempio, possiamo osservare come il verbo indossare costituisca la realizzazione dello scopo per cui la camicia è pensata. Potremmo quindi parafrasare con:  <i>Federico realizza lo scopo della camicia, ovvero fa con la camicia quel che si deve fare con la camicia.</i></p> <p><b>Real<sub>1</sub></b>(promessa) = mantenere [DET ~]  <i>Nicola ha mantenuto la sua promessa</i></p>

				<p>(= <i>Nicola ha realizzato lo scopo della promessa</i>)</p> <p><i>Promessa</i> ha tre attanti: (1) chi promette, (2) ciò che viene promesso, (3) la persona a cui si promette. Ecco perché <i>Real</i><sub>1</sub>.</p> <p>Si confronti con:</p> <p><b>Oper</b><sub>1</sub>(promessa) = fare [una ~]</p>
53	<i>Facto</i> <sub>0/i</sub>	Verbo di realizzazione il cui I attante è la base	<p>Il valore della funzione è un verbo di realizzazione che significa in sintesi “[la base] realizza la sua funzione/si realizza”.</p> <p>La sua struttura attanziale è analoga a quella di <i>Func</i> (possiamo dire che <i>Fact</i> ne sia l’equivalente, ma con il senso di realizzazione):</p> <ul style="list-style-type: none"> <li>- il <b>primo attante</b> sintattico profondo è la base,</li> <li>- l’eventuale secondo attante è l’attante semantico <i>i</i> della base.</li> </ul>	<p><i>Facto</i>(coltello) = tagliare</p> <p><i>Facto</i>(sogno) = [~] realizzarsi</p> <p><i>Facto</i>(lampadina) = [~] illuminare</p>
54	<i>Labreal</i> <sub>ij</sub>	Verbo di realizzazione il cui III attante è la base	<p>Il valore della funzione è un verbo di realizzazione (come <i>Real</i> e <i>Func</i>)</p> <p>La sua struttura attanziale è analoga a quella di <i>Labor</i> (di cui rappresenta l’equivalente + senso di realizzazione):</p> <ul style="list-style-type: none"> <li>- il suo primo attante sintattico profondo (o soggetto) è l’attante semantico <i>i</i> della base</li> <li>- il secondo attante (o complemento) è l’attante semantico <i>j</i> della base</li> <li>- il <b>terzo attante è la base.</b></li> </ul>	<p><b>Labreal</b><sub>12</sub>(frullatore) = tritare [Y con un ~]</p> <p><i>Il cuoco trita la carne con il frullatore</i></p> <p>Frullatore è una parola che designa un “artefatto che serve affinché X [=attante 1] triti Y [=attante 2]”</p>
55	<i>Involv</i>		<p>Il valore della funzione è un verbo il cui senso è “[la base] coinvolge, riguarda Y”, dove Y è un elemento esterno che è coinvolto nella situazione descritta dalla base.</p> <p>La struttura attanziale è la seguente:</p>	<p>EN: <i>Involv</i>(light) = [~] floods [Y]</p> <p><i>The light flooded the room</i></p> <p>EN: <i>Involv</i>(snowstorm) = to catch [Y LOC<sub>in</sub> N]</p>

			- la base è il primo attante - Y è il secondo attante.	<i>The snowstorm caught John in the middle of nowhere</i> <sup>53</sup>
56	Manif		Il valore della funzione è un verbo che significa “[la base] si manifesta in Y”, dove Y è un elemento esterno in cui la base manifesta i suoi segni. La struttura attanziale è la seguente: - la base è il primo attante - Y è il secondo attante.	Manif( <i>amazement</i> ) = <i>lurks [in Andrew's eyes]</i> Manif( <i>joy</i> ) = <i>explodes [in them]</i> Manif( <i>scorn</i> ) = <i>is dripping [from every word]</i> <sup>54</sup>
57	Prox		Funzione lessicale che apporta il senso “stare per...”. Non ha struttura attanziale, fa sempre parte di FL complesse assieme a FL verbali.	ProxOper <sub>1</sub> ( <i>crisi</i> ) = essere sull’orlo di [DET ~]
58	Prepar <sub>i</sub>		Funzione che apporta il senso “preparare qualcosa/prepararsi per l’uso o il funzionamento. Non ha struttura attanziale, fa sempre parte di FL complesse assieme a FL verbali. Esiste un evidente nesso tra Prox e Prepar: Prepar equivale a CausProx(L), con in più il senso di intenzione, poiché l’azione di preparazione è svolta di proposito, mentre Prox può indicare azioni involontarie.	PreparFact <sub>0</sub> ( <i>software</i> ) = caricarsi Prepar1Real2( <i>plane</i> ) = <i>[to] board [a ~]</i> <sup>55</sup>
59	Degrad		Il valore della funzione è un verbo che esprime un deterioramento permanente della base.	Degrad( <i>latte</i> ) = [~] inacidirsi EN: Degrad( <i>clothes</i> ) = [~] wear off
60	Son	Suono	Il valore della funzione è un verbo che designa il suono	Son( <i>telefono</i> ) = [~] squillare

<sup>53</sup> Esempi tratti da Mel’čuk (1996:69)

<sup>54</sup> Esempi tratti da Mel’čuk (1996:70)

<sup>55</sup> Esempio tratto da Mel’čuk (1996:70)

			convenzionale prodotto dalla base.	Son(gatto) = {[~] miagolare; [~] fare le fusa, ...} EN: Son(cricket) = [~] to chirp
61	Obstr <sub>i</sub>		Verbo che esprime il senso “[la base] funziona con difficoltà”.	EN: Obstr(negotiations) = [~] are stalled EN: Obstr(economy) = [~] stagnates <sup>56</sup>
62	Stop <sub>i</sub>		Verbo che significa “[la base] smette di funzionare”.	Stop(voce) = [~] spezzarsi
63	Excess <sub>i</sub>		Verbo che significa “[la base] funziona in modo anormale ed eccessivo”.	EN: Excess(engine) = [~] races <sup>57</sup>
64	Sympt <sub>ij</sub>		<p>Il valore della funzione è un verbo che denota una reazione fisica sintomo di uno stato emotivo o fisico particolare (indicato dalla base).</p> <p>In genere questa FL compare all'interno di FL complesse assieme alle funzioni Obstr, Stop, Excess.</p> <p>La struttura attanziale è la seguente:</p> <ul style="list-style-type: none"> <li>- l'attante 1 del verbo è la parte del corpo interessata dal sintomo</li> <li>- l'attante 2 è la persona che presenta il sintomo</li> <li>- l'attante 3 è lo stato emotivo o fisico che è causa del sintomo.</li> </ul>	<p>EN: Obstr(breath) - Sympt<sub>23</sub>(anger) = [N] chokes [with ~]</p> <p>EN: Obstr(speech)-Sympt<sub>23</sub>(anger) = [N] sputters [with ~]</p> <p>Stop(speech) - Sympt<sub>1</sub>(amazement) = //[N] is dumbstruck</p> <p>Excess<sup>motor</sup> (mouth)<sup>58</sup> - Sympt<sub>213</sub>(amazement) = [N] opens [N's] mouth wide [with amazement]<sup>59</sup></p>

<sup>56</sup> Esempi tratti da Mel'čuk (1996:71).

<sup>57</sup> Esempio tratto da Mel'čuk (1996:72).

<sup>58</sup> L'indice motor in apice specifica un aspetto del significato della base, ossia il suo movimento. Excess<sup>motor</sup> indica un movimento anormale.

<sup>59</sup> Esempi tratti da Mel'čuk (1996:72).

## Appendice 3: Scheda del lessema DIRITTO<sub>2</sub>

La presente scheda è stata elaborata secondo i criteri della lessicografia esplicativa-combinatoria e utilizzando come parziale riferimento il DiCoLiLex, database elettronico delle collocazioni del francese attualmente in corso di realizzazione all'Université de Montréal (cfr. Marengo e Robichaud, approvato). La scheda non riproduce la struttura classica di una voce del Dizionario esplicativo-combinatorio, ma si concentra unicamente sugli aspetti semantici e relativi alle proprietà combinatorie. Le informazioni relative alla struttura di attanti sintattici e alla diatesi sono state omesse.

La scheda contiene informazioni esplicite sulla struttura di attanti semantici del lessema DIRITTO e un elenco di funzioni lessicali, con il collocato (o i collocati) corrispondente al valore di ciascuna funzione lessicale, sul modello del DEC.

Occorre precisare che il lessema DIRITTO è stato qui analizzato unicamente nella sua accezione “facoltà”, tutte le informazioni relative all’accezione “insieme di norme” sono state omesse.

Dal momento che tutti i dizionari consultati (De Mauro, Treccani, Tiberii, Urzì) riportano l’accezione presa in esame come seconda accezione, accanto al nome del lessema si riporta l’indice <sub>2</sub>, a segnalare che si tratta della seconda accezione del termine: DIRITTO<sub>2</sub>.

La scheda è articolata in tre sezioni: 1. Definizioni (tratte da due dizionari di lingua italiana); 2. Forma proposizionale (cfr. cap.1 sez. 1.6.1.1); 3. Tabella delle funzioni lessicali.

### DIRITTO<sub>2</sub>

#### 1. Definizioni

- facoltà, giuridicamente tutelata, di operare proprie scelte e di esigere che altri mettano in atto o si astengano da un determinato comportamento (De Mauro, 2016)
- Facoltà o pretesa, tutelata dalla legge, di un determinato comportamento attivo o omissivo da parte di altri (Dizionario Treccani)

#### 2. Forma proposizionale

*Diritto della persona X a Y, conferito dall'autorità Z, di fronte a W.*

Esempio: diritto del lavoratore (X) al riposo settimanale (Y), garantito dallo Statuto dei Lavoratori (Z) e fatto valere di fronte al datore di lavoro (W).

Il primo attante semantico (X) è il titolare del diritto, il secondo attante semantico (Y) è l’oggetto del diritto, il terzo attante semantico (Z) è l’autorità o la legge che conferisce il diritto, e il quarto attante semantico (W) è la persona o l’entità di fronte a cui il diritto viene fatto valere.

### 3. Tabella delle Funzioni lessicali standard (semplici e complesse)

Legenda:

DET = determinante (articolo, aggettivo possessivo/dimostrativo)

~ = parola-chiave

X = primo attante semantico della parola-chiave

Y = secondo attante semantico della parola-chiave

Z = terzo attante semantico della parola-chiave

W = quarto attante semantico della parola-chiave

Nella colonna **Collocazione** le parentesi quadre racchiudono l'attante semantico della base che costituisce il soggetto del verbo.

NB: alcune funzioni lessicali riportano indici numerici.

Gli indici in cifra araba in pedice (es: Oper<sub>1</sub>) designano gli attanti della base, mentre gli indici in cifre romane in apice (es: Real<sub>4</sub><sup>II</sup>) servono a indicare diverse sfumature di senso interne alla specifica funzione lessicale. Questo perché a una stessa funzione lessicale possono corrispondere verbi che presentano sfumature di senso diverse.

La colonna **Note** riporta alcune precisazioni relative alle varie funzioni lessicali, senza alcuna pretesa di esaustività: la presente scheda è infatti ispirata alla struttura delle voci del DEC, in cui i collocati sono elencati in corrispondenza della relativa funzione lessicale, ma le funzioni lessicali in sé non sono descritte. Ricordiamo che nel presente lavoro le caratteristiche generali delle varie funzioni lessicali sono già illustrate in Appendice 2.

Funzione Lessicale	Collocazione	Note
S <sub>1</sub>	beneficiario di DET ~ titolare di DET ~ avente ~	
S <sub>2</sub>	oggetto di ~	
Ver	~ inviolabile; ~ legittimo	
Ver + Adv <sub>1</sub>	a buon ~	
Func <sub>0</sub>	DET ~ esiste	
Func <sub>1</sub>	a [X] spetta DET ~	
FinFunc <sub>0</sub>	~ venire meno	
Oper <sub>1</sub>	[X] avere DET ~ [X] godere di DET ~ [X] beneficiare di DET ~ <sup>60</sup>	In questo caso Y dipende da DIRITTO, non è un attante di Oper <sub>1</sub>

---

<sup>60</sup> *beneficiare* di un diritto può assumere un senso diverso a seconda del contesto: in alcuni casi può corrispondere alla funzione lessicale Oper<sub>1</sub> (=avere un diritto), mentre in altri casi può assumere una sfumatura di senso vicina a *esercitare* (Fact<sub>1</sub>), nel qual caso corrisponde alla funzione lessicale Fact<sub>1</sub>. Cfr. sez. 4.5.

IncepOper <sub>1</sub>	[X] ottenere DET ~ [X] acquisire DET ~	
FinOper <sub>1</sub>	[X] perdere DET ~ [X] decadere da DET ~	
Oper <sub>2</sub>	[Y] essere (ART) ~ [Y] costituire ART ~ [Y] rappresentare ART ~ [Y] rientrare fra DET <sub>PL</sub> ~ <sub>PL</sub>	Esempi: “Votare è (un) mio diritto”; “La difesa rientra fra i diritti dell’imputato”
Oper <sub>3</sub>	[Z] stabilire DET ~ [Z] sancire DET ~ [Z] garantire DET ~	Z corrisponde tipicamente a un’istituzione, a una legge, <sup>61</sup> ecc.
Caus <sub>3</sub> Func <sub>1</sub>	[Z] concedere DET ~ a X [Z] dare DET ~ a X	
Real <sub>1</sub>	[X] esercitare DET ~ [X] far valere DET ~ [X] avvalersi di DET ~ [X] beneficiare di DET ~	
AntiReal <sub>1</sub>	[X] rinunciare a DET ~	“Rinunciare a un diritto” non significa perderlo, ma scegliere di non esercitarlo in un momento definito.
AntiBonReal <sub>1</sub>	[X] abusare di DET ~	
Real <sub>3</sub>	[Z] tutelare DET ~ [Z] garantire DET ~	In frasi come “la legge tutela il diritto allo studio”, non si tratta tanto del senso “proteggere” (=Perm <sub>3</sub> Func <sub>0</sub> ) ma di un senso di realizzazione: la legge “fa ciò che deve fare” (=Real <sub>3</sub> ). Anche “Garantire” è un verbo ambiguo, perché può esprimere sia un Oper <sub>3</sub> che un senso di realizzazione (=Real <sub>3</sub> ).
Real <sub>4</sub> <sup>l</sup>	[W] riconoscere DET ~	L’indice / in cifra romana in apice indica una diversa sfumatura di senso: “Riconoscere un diritto” e “rispettare un diritto” corrispondono entrambi a Real <sub>4</sub> , ma esiste una differenza significativa tra i due verbi: riconoscere un diritto è semplicemente riconoscerne l’esistenza, mentre rispettarlo corrisponde a un grado maggiore di “realizzazione”. Per marcare questa

<sup>61</sup> nel vocabolario Treccani *sancire* = stabilire mediante una legge: la Costituzione italiana sancisce il diritto al lavoro di tutti i cittadini.

		differenza si è assegnato l'indice I a "riconoscere" e l'indice II a "rispettare" (cfr. sotto).
Real <sub>4</sub> <sup>II</sup>	[W] rispettare DET ~	In questo caso l'indice II in cifre romane in apice serve a segnalare un maggiore "grado di realizzazione" rispetto a Real <sub>4</sub> <sup>I</sup> : rispettare un diritto significa necessariamente averlo prima riconosciuto.
Real <sub>41</sub>	[W] riconoscere a X il ~	In questo caso i due indici in cifra araba corrispondono agli attanti semantici della base che devono necessariamente essere espressi in forma di attanti sintattici profondi (argomenti obbligatori) del verbo. La differenza tra Real <sub>4</sub> <sup>I</sup> e Real <sub>41</sub> è che nel primo caso l'espressione dell'attante 1 della base non è obbligatoria, mentre nel secondo sì: "riconoscere" può realizzarsi come un verbo a 2 argomenti oppure a 3 argomenti. Dal momento solo "riconoscere" può ammettere una costruzione a tre attanti, si ritiene di poter omettere gli indici in cifre romane che differenziano Real <sub>4</sub> <sup>I</sup> e Real <sub>4</sub> <sup>II</sup> .
AntiReal <sub>4</sub> <sup>I</sup>	[W] non riconoscere DET ~ [W] negare DET ~	
AntiReal <sub>4</sub> <sup>II</sup>	[W] non rispettare DET ~; [W] violare DET ~	
AntiReal <sub>4</sub> <sup>III</sup>	[W] ledere DET ~; [W] intaccare DET ~ [W] pregiudicare DET ~	L'indice III in cifre romane in apice serve a segnalare una sfumatura di senso diversa sia da AntiReal <sub>4</sub> <sup>I</sup> che da AntiReal <sub>4</sub> <sup>II</sup> : il senso di "intaccare i diritti" o "ledere i diritti" implica infatti un comportamento lesivo a lungo termine, e non un danno "puntuale" come nel caso di una violazione, che può costituire un singolo episodio.
AntiReal <sub>41</sub>	[W] negare DET ~ a X	
SoAntiReal <sub>4</sub> <sup>I</sup>	negazione di DET ~	
SoAntiReal <sub>4</sub> <sup>II</sup>	violazione di DET ~	
MagnAntiReal <sub>3</sub>	[W] farsi beffe di DET ~	
SoReal <sub>4</sub> <sup>II</sup>	rispetto di DET ~	
AdvReal <sub>4</sub> <sup>II</sup>	nel rispetto di DET ~	

LiquOper <sub>1</sub>	privare X di DET ~	Nella RSintP, il primo attante è l'agente, il secondo attante è X e il terzo attante è ~. Per effetto della funzione lessicale causativa, la funzione verbale con tre attanti viene "declassata" a una funzione a due attanti (OPER), poiché l'agente causatore è «esterno».
MagnLiquFunc <sub>0</sub>	annientare DET ~	
A <sub>2</sub> LiquOper <sub>1</sub>	privato di DET ~	È A <sub>2</sub> perché è l'aggettivo che definisce il secondo attante sintattico profondo di LiquOper <sub>1</sub> .
Caus <sub>1</sub> Func <sub>0</sub>	[X] riservarsi DET ~ [X] conquistare DET ~	Qui X è già l'attante 1 di Caus, non può fare anche da attante di Func, quindi è Func <sub>0</sub> (e il motivo per cui è Func (e non OPER) è la "regressione degli attanti" in presenza di funzioni lessicali di causa.
PermFunc <sub>0</sub>	difendere DET ~ tutelare DET ~ proteggere DET ~	
S <sub>0</sub> PermFunc <sub>0</sub>	difesa di DET ~ Tutela di DET ~ salvaguardia di DET ~	
S <sub>1</sub> PermFunc <sub>0</sub>	difensore di DET ~	

### Funzioni lessicali non standard

<i>Cercare di</i> Caus <sub>1</sub> Func <sub>0</sub>	rivendicare DET ~	
<i>Malgrado un'opposizione da parte di Z o W,</i> Caus <sub>1</sub> Func <sub>0</sub>	arrogarsi DET ~	
<i>Fare qualcosa che potenzialmente</i> Liqu <sub>30</sub> per <sub>1</sub>	[Z] minacciare DET ~	