


Saggi
Essays

Risorse linguistiche per la generazione automatica della lingua dei segni naturali

Linguistic Resources for Automatic Natural Sign Language Generation

Ritamaria Bucciarelli

Università degli Studi di Siena • ritamaria.bucciarelli@unisi.it

Valentina Falco

Ordine forense di Salerno • valentinafalco@pec.ordineforense.salerno.it

Giulia Savarese

Università degli Studi di Salerno • gsavarese@unisa.it

ABSTRACT

The paper focuses on “Technology Training”: blended teaching, sup-brought by technological tools. Punto strength of the reviews on the subject are the descriptions of the projects and research products, which highlight the cognitive and language teaching technology tools potential used the Center of Didactics of Languages of the “Ca’Foscari” University of Venice. The starting point are the glottotechnologies of a corpus, intended as a tool for analysis, translation and text formulation of the language, both for the use of the common language for a certain type of Fachsprache. The description of the various methods for acquiring or extending skills for the drafting of specialist texts and for developing large linguistic resources such as NLG software is provided. Attention will then be given to “neutral” language applications and resources (dictionaries, morphology, sentence structure and transformation grammars), which can be used both to analyze and automatically generate INLG 2017, texts as Acro-Word (2014), that is a multifunctional software, that substitute the paper sheet and that is supported by man produces different types of text in paraphrases and free.

Il paper si sofferma sulla tematica della “Technology Training”: blended teaching supportati da Technological tools¹. Punto di forza delle disamine saranno le descrizioni dei progetti e prodotti di ricerca, che evidenziano le potenzialità cognitive e glottodidattiche degli strumenti tecnologici utilizzati dal Centro di Didattica delle Lingue dell’Università “Ca’ Foscari” di Venezia. Il punto di partenza sono le glottotecnologie di un corpus, intese come strumento di analisi, di traduzione e di formulazione testo della lingua, sia per l’uso della lingua comune che per un determinato tipo di Fachsprache. Si giunge alla descrizione delle diverse modalità per acquisire o ampliare le competenze per la stesura di testi specialisti e a sviluppare risorse linguistiche di grande copertura quali software NLG reali. Si pone, poi, attenzione alle applicazioni e alle risorse linguistiche “neutrali” (dizionari, morfologia, struttura delle frasi e grammatiche di trasformazione), che possono essere utilizzate sia per analizzare e generare automaticamente i testi INLG 2017, quale l’Acro-Word (2014), un software plurifunzionale, che sostituisce il foglio cartaceo e che, supportato dall’uomo, produce diverse tipologie testuali in parafrasi e a testo libero.

KEYWORDS

Computational Linguistics; Computer Technology; Knowledge of Online Grammars; Glottotechnologies; Blended Teaching.

Linguistica Computazionale; Informatica; Conoscenza delle Grammatiche Online; Glottotecnologie; Blended Teaching.

1 Attività di ricerca – Università di Venezia “Ca’ Foscari”, insegnamento “Corsi ufficiale Italiano scritto[FTO302]” – <https://moodle.unive.it/course/info.php?id=132>, un progetto d’ateneo dir. Prof. Paolo Balboni (2003-2017), in collaborazione con altri Centri di ricerca e, in particolare, con il Laboratoire d’Automatique et Linguistique (C. N. R. S. – Università Paris 7), in cui sono stati messi a punto nuovi metodi per l’indagine linguistica e per una moderna glottodidattica ispirata da M. Silberstein.

1. Introduzione

Nella galassia del linguaggio, la comunicazione sintetica, fatta di frasi fisse o parafrasi di frasi fisse (testi precostituiti), sta diventando sempre più comune. Questo tipo di comunicazione richiede risorse linguistiche in grado di garantire comunicazioni specialistiche, cioè modelli precisi ottemperanti le tipologie, ma soprattutto redatti con terminologie appartenenti a determinati ambiti professionali, che riproducano una semantica breve, ma soprattutto che rispecchino la grammatica formale e l'ordine frastico Bucciarelli (2015). Nel corso di un ultradecennale lavoro di sperimentazione condotto presso Atenei italiani, il modello di riferimento pilota e oggetto dei miei studi è stato quello appreso presso il *Dipartimento di Scienze della Comunicazione* dell'Università di Salerno, in collaborazione con altri Centri di ricerca e, in particolare, con il *Laboratoire d'Automatique et Linguistique* (C. N. R. S. – Università Paris 7). Così sono stati messi a punto nuovi metodi per l'indagine linguistica, basati essenzialmente sulla costruzione di lessici sintattici che, giovandosi delle opportunità offerte dall'elaborazione informatica dei dati, mirano a una descrizione, la più esaustiva e formalizzata possibile, di una data lingua. Le ricerche fanno parte del progetto *ALTI* ricerca nazionale (PRIN) 2005–2008². La sede di ricerca dell'Università di Salerno fa ricerca sugli diversi aspetti lessicali e sintattici delle sequenze cosiddette polirematiche presenti in italiano – inglese e la metodologia è quella lessico-grammaticale, che fu originariamente introdotta nella comunità scientifica europea da Maurice Gross all'Università Parigi 7 dopo il 1960. Lo scopo principale del lessico-grammaticale è di descrivere dettagliatamente tutti i meccanismi combinatori indissolubilmente legati alle concrete entrate lessicali. Uno dei risultati più importanti della descrizione lessico-grammaticale è la costruzione di *software* e *lingware* orientati all'analisi testuale automatica. La ricerca si basa sul reperimento e sull'analisi di forme linguistiche, sia semplici sia composte (polirematiche), all'interno di ampi corpus testuali, le stesse inserite, poi, in base dati ed in conformità con gli standard europei *RELEX*. I data base elettronici sono stati poi integrati ed utilizzati in software d'analisi testuale automatica ed utilizzati per il *tagging* morfo-sintattico di testi in formato elettronico (De Bueris & Elia 2008, pp.11-12), sostengono:

In questa fase sono state costruite grammatiche locali sotto forma di grafi e trasduttori a stati finiti. Una grammatica locale è una grammatica elettronica costruita per descrivere solo una specifica caratteristica di un data lingua. Nel nostro caso, le grammatiche locali sono applicate ai testi per recuperare informazioni lessicali, morfologiche e sintattiche. Scopo di questo progetto di ricerca è quindi dimostrare come, partendo dall'analisi delle parole, è possibile realizzare l'analisi automatica di combinazioni di parole, ovvero l'analisi automatica di interi sintagmi e frasi complete. La nostra unità di Ricerca ha realizzato un lessico elettronico terminologico polirematico nel dominio dell'economia che consta di circa 30.000 entrate in forma flessa, equivalenti a circa 15.000 forme canoniche. Di qui l'esperienza per il trattamento automatico, quindi analisi delle frasi fisse effettuata con Nooj e lo studio dei nuovi software.

- 2 Un programma per l'elaborazione del linguaggio naturale sviluppato da Max Silberztein (a partire dal 2007), che permette la costruzione dizionari e di grammatiche elettroniche e la loro applicazione a corpora di grandi dimensioni (cfr. per approfondimenti www.nooj4nlp.net).

Di qui l’idea di produrre non solo lessici elettronici, descrittori lessicali di forme sintattiche, morfologiche ed ortografiche, ma nuove applicazioni per il trattamento automatico del linguaggio naturale.

2. Ipotesi operativa

Il primo approccio è *Type-Race* (Bucciarelli, 2015)³, la prima risorsa linguistica per la elaborazione automatica del linguaggio naturale, da me ideata e progettata con il team del Dipartimento di Scienze della comunicazione, “Polo DSN” dell’Università degli Studi di Salerno. L’elaboratore del linguaggio naturale è capace di generare automaticamente morfemi o parti di parole. Il *Type-Race* in una prima versione prototipale è costituita da diversi momenti interagenti, perché si intuisce che, potenziando gli ambiti delle conoscenze linguistiche, migliora qualitativamente l’output informatico. L’ipotesi esplorativa di linguistica computazionale, in fase sperimentale è diretta ad analisi del sintagma e ancor più a sintesi di elementi morfemici e lessamici, atti a produrre tecniche di apprendimento cognitivo per non complessità strutturale e semantica di testi di carattere normativo, quali quelli presenti in circolari, decreti, leggi dello stato. Esso prevede la realizzazione di un *parser* morfologico, sintattico-semantico, che è in grado di operare scelte strutturali sulla base di codici personalizzati, quali la posizione di un certo costituente, nonché di strutture fisse, formule utilizzate in varie parti del testo in analisi, allo scopo di produrre tassonomie di polirematiche e di elementi lessemici ridotti in acronimi, che servono a compilare una base dati, che deve automaticamente riprodurre la frase no-composizionale e fissa all’atto del richiamo. Questa prima fase viene denominata di “manufacturer and generator acronyms”. La seconda fase inizia con il richiamo e l’elaborazione dell’informazione da parte di un operatore, che deve produrre un testo e che rappresenta il primo tentativo di comporre un testo formato da parafrasi di frasi in cui l’operatore digitale riformula il linguaggio e cioè aggiunge alla unità di significato determinanti frasi o parti di frasi libere (vedi Fig. 1).



Fig. 1. Type-Race

3 Type -Race 2000 © R. Bucciarelli; M. Renda. All. Rights reserved. N50061363717.

3. Acro-Word: Text Constructor

Nel 2014 nasce l'*Acro-Word* (Bucciarelli, 2015)⁴, una nuova risorsa linguistica per la riproduzione e la gestione automatica dei linguaggi NLG, ma soprattutto per i linguaggi della *new-economy*. Il software ha come scopo di supportare la produzione dei testi o di parti testuali. Il text constructor è capace di generare di produrre testi dei linguaggi NLG e on più solo in parafrasi (invarianza morfemica e sinonimia), ma test composti da osmosi di frasi libere e frasi fisse. Per la prima volta è previsto l'intervento di un linguista digitale, che sulla base delle sue conoscenze, formula, riformula traduce e trasforma un testo digitale, cioè un software per l'analisi testuale automatica e di un costruttore testi per la comunicazione dei linguaggi NLG. Prima di giungere alla costruzione di un testo digitale, è opportuno esaminare il corpus in omnia parte, perché ogni parte ha una funzione autonoma, ma interagente sul modello lessico-grammatica, che, grazie alla linguistica computazionale, valida il metodo empirico delle sue ricerche e crea software per l'analisi testuale automatica, che si servono di *software* applicativi come dizionari elettronici e pertanto composto da momenti interagenti.

3.1. Fase A: "Manufacturer and generator acronyms"

Nella fase A si focalizza l'unità polirematica, perché si intuisce che per avere una risorsa linguistica di parti testuali fisse è necessario partire dalla produzione di tassonomie di frasi fisse, che, come le definisce De Mauro (1999, p.11), esse:

sono un gruppo di parole che ha un significato unitario e non deducibile dalle parole che lo compongono, ma dalla semantica unitaria della unità.

La ricerca parte dall'esperienza acquisita dalla partecipazione al progetto PRIN dell'Università degli Studi di Salerno, presso il "Polo DSN", per la produzione del dizionario elettronico delle unità polirematiche *DELACF* dell'italiano e sugli approfondimenti degli studi sul metodo di formalizzazione del linguaggio naturale, del lessico-grammatica elaborato da Maurice Gross. Infatti, dall'approfondimento sugli studi di frasi idiomatiche, polirematiche e gruppi morfemici misti, si intuisce, così, che un linguaggio fisso deve essere costruiti per ambiti specialistici composte da strutture fisse quali linguaggi specialistici. Le lingue *LSP* non sono usate per usi specialistici e, pertanto, non lingue autonome, divise dal linguaggio ordinario, ma incluse e osmoticamente condivise in "accezioni specifiche" (De Mauro, 1999, pp. 410-411). A tale proposito, Longobardi, (2011, p. 129) definisce:

il ruolo dell'avverbio composto nella descrizione dei linguaggi specialistici e chiarisce sul punto che i ricercatori e linguisti associano le proprietà semantiche e sintattiche alla elaborazione automatica dei testi, ma non si occupano abbastanza per il recupero e la selezione della informazione.

4 *Acro-Word* 2014 © R. Bucciarelli; P. Villari, F.; Terrone, M. Terrone; F. Santoro; R. Marco-ne. All. Rights reserved. N50061363717.

Infatti, nella costruzione delle schede terminologiche investono di rado sul funzionamento sintattico-semantic. È proprio questa asserzioni a farci giungere all'idea che, associando le proprietà sintattiche alla elaborazione automatica dei testi per il recupero dati, allo studio del funzionamento sintattico semantico, si può ottenere un recupero totale delle informazioni. La ricerca si orienta in questa prima fase sul reperimento e sull'analisi di forme linguistiche, composte (polirematiche) che all'interno di ampi corpus testuali. Si passa alla progettazione del prodotto *A.W.* e inizia la fase di compilazione delle unità semantiche; vengono ridotte in acronimi e inserite, poi, in base dati; successivamente, si completa così l'indicizzazione e la *tokenizzazione*, per poi eseguire il *matching* delle unità semantiche contenute nelle tassonomie e le entrate catalogate. Il *matching* dà il dizionario elettronico delle unità semantiche, in quanto unità di significazione di un acronimo. Le entrate sono elencate in ordine alfabetico e contengono informazioni di tipo morfo-grammaticale e sono suddivise in base alla loro caratteristica di unità di significato autonome. È proprio nell'analisi di queste forme il riferimento al modello teorico e metodologico lessico-grammaticale, individuato da Maurice Gross (1988, 1989, 1990) ed applicato alla lingua italiana da D'Agostino e Elia (1998), delle polirematiche, che si addivene ad un recupero automatico delle informazioni di carattere semantico e si ottiene un dizionario elettronico che, all'interno del software di analisi testuale, svolge, il ruolo di motore linguistico con cui effettuare tutte le operazioni di *matching*. Pertanto, si costruisce il *software A. W* che in ogni parte del prodotto singolarmente ha le sue funzioni, che sono:

- A - di consultazioni ed analisi morfo-sintattico del dizionario elettronico. In questa fase è un analizzatore testuale automatico lessico-grammatica di tipo morfosintattico che usa motori linguistici, servendosi di base dati lessicali incorporati all'interno di una *shell* pacchettizzata e composta da *software* modulari sul modello *NOOJ*;
- B - di lettura automatica di un testo, completa di indicizzazione e *tokenizzazione*, per eseguire il *matching* tra le parole contenute dal testo e le entrate catalogate e classificate nei dizionari elettronici. Il risultato del *matching* è la creazione di dizionari elettronici del testo analizzato; in questi dizionari, le entrate sono elencate in ordine alfabetico, corredate da informazioni di tipo morfo-grammaticale e suddivise in base alla loro caratteristica di unità di significato autonomo;
- C - in una seconda fase dell'analisi, di lettura all'interno del testo, effettuando specifiche ricerche visualizzabili sotto forma di concordanze, nonché di effettuare la localizzazione di *pattern* sintattici, la disambiguazione ed il *parsing* del testo;
- D - in una terza ed ultima fase, di importare specifici file realizzati in forma di tabella con *Microsoft Office Access*®;
- E - infine, di inserimento di immagini, filmati documenti, che sono utilizzabili e visionabili all'occorrenza (vedi Fig. 2).

550	\$(un>	voto all'unanimità	voto all'unanimità	voto all'unanimità
551	\$(p>	voto ponderato	voto ponderato	voto ponderato
552	\$(to>	world trade organization	world trade organization	world trade organization
553	\$(y>	youth for europe	youth for europe	youth for europe
554	\$(z>	zona di libero scambio	zona di libero scambio	zona di libero scambio
555	\$(zr>	zona rifugio	zona rifugio	zona rifugio
556	\$(zptn>	Chi non dà valore al centesimo, non meri	Wer den Pfennig nicht ehrt, ist den taler	Who don't take in care ce
557	\$(s>	Lingua italiana dei segni	Italienische Gebärdensprache	Italian sign language
558	\$(DFB>	Dattilologia Fonetica Bimanuale	Bimanuelle Fingeralphabet Phonetik	Bimanual finger spelling Ph
559	\$(RAI>	Telegiornale con sottotitolazione	Nachrichten mit Untertiteln	News with subtitles
560	\$(FSB>	Corpus LFSB Le puzzle		
561	\$(
567	\$(ANIMAV>	Le basta solo un po' d'aria nuova. Se mi g		
568	\$(AMICO>	È facile allontanarsi sai. Se come te anch		
569	\$(IDGS>	Progetto di Hamburg	Entwurf Amburg	Project Hamburg
571	\$(AMUSCO>	La vita in rosa		
573	\$(SENTO>	Sordo, mi chiamano Sordo!		

Fig. 2. Software A. W – Descrizione della codifica e riproduzioni di Frasi fisse

3.2. Fase 2: “Man builder text”

Il costruttore del testo, come dicono D’Agostino e Elia (1998), è un professionista della scrittura che redige un testo digitale in L_1 , in linguaggio naturale in codice della ricevenza del messaggio e quindi trasmette in L_2 , opera sui tecnicismi della lingua da trasmettere, partendo dall’operatore verbale di L_2 producendo analisi delle manipolazioni trasformazionali e sostituzionali. Si ricordi, a tal proposito, la visione filosofica kantiana, cioè che la visione di un’immagine cognitiva di ogni uomo si sviluppa traducendosi in pensiero secondo il suo modo di interagire con i contenuti e le reali capacità comunicative. In questa fase 2, la ricerca passa dalla focalizzazione unità polirematica alla frase semplice. Il modello di riferimento in cui si identifica il costruttore testo di *Acro-Word* è nel metodo di formalizzazione del linguaggio naturale, il lessico-grammatica elaborato da Maurice Gross, che si basa principalmente sulle conclusioni strutturaliste, trasformazionali e distribuzionali di Harris. Da quest’ultimo, Gross (1995) riprende soprattutto i concetti di trasformazione linguistica e di frase nucleare (che diviene frase semplice nel lessico-grammatica), riportandoli all’interno del quadro di una grammatica formale delle lingue naturali secondo cui il lessico è solo l’insieme di valori terminali da associare a sequenze ordinate in base a regole e principi combinatori autonomi. Il modello teorico di riferimento è rappresentato dalla grammatica di D’Agostino e Elia (1998) ispirato “a operatori e argomenti” di Z. S. Harris (1970). L’operatore digitale, sulla base delle sue conoscenze, guida la stesura del testo da redigere e cioè sceglie di descrive i foni, filtrando i toni; di costruire degli assi sintagmatici; di produrre frasi libere; di riproporre le frasi fisse, secondo la tipologia testuale del testo in produzione; di applicare la descrizione della morfologia e sintassi della lingua d’uso (Elia et al., 1985).

Volendo procedere a una classificazione tassonomica delle frasi possibili nell’italiano scritto al fine di ottenere una traduzione corretta o un testo specialistico, è opportuno chiarire l’importanza degli elementi costruttivi, mediante il metodo di ricerca e di sperimentazione del *L. G. L. I.* (Elia et al., 1985).

Sulla base di queste premesse necessita:

- selezionare le frasi sintetiche ed i livelli di frequenza;
- valutare e comparare le manipolazioni di trasformazione e di sostituzione;
- conoscere le teorie degli assi in L_2 (Arcuri & D’Agostino, 1982).

A tale proposito, Bucciarelli (2016, p. 9) afferma:

All’aridità e staticità dei segni intendesi sostituire la mobilità dell’intelligenza umana, che, alimentata dallo spirito, vivacizza l’apprendi-

mento, rendendolo esaustivo ed esplicativo dei segni, talora forieri di dubbi, perché non sempre comprensibili: laddove non arriva la macchina, interviene l'uomo con la sua sensibilità ed il suo pensiero. L'esponente ritiene che i segni diano verità universali ed ineccepibili, ma solo la mano e la mente umana sappiano trovare la strada idonea alla comprensione, all'elaborazione e alla interiorizzazione dello scibile. Con la giusta miscela di scienza informatica e umana sapienza, con un'adeguata manipolazione e contaminazione della evoluzione e della tecnica con le capacità della persona, si riesce a suscitare la reattività dell'altro, a cristallizzare la conoscenza in un'unica ed amena sintesi; soprattutto si può pervenire a risultati forse meno moderni ma senz'altro più sicuri per tutti. L'esponente, pioniera di una nuova strategia lessicale, di una sperimentazione mista di scienza e di umano sentire, si propone di perseguire obiettivi non ancora raggiunti e di elevare il potere di conoscenza anche ai meno dotati. Auspica che il suo progetto proceda senza personalismi, bensì venga accolto da una vasta platea con unanime consenso. È dell'idea che, solo con la partecipazione e con la condivisione su larga scala, il suo impegno ed il suo operato possano avere un riscontro positivo, scuotere l'interesse di molti, soprattutto arrecare beneficio anche a coloro, che ancora ignorano le conoscenze basilari dei segni.

La macchina diventa un supporto un *container* nelle menti specialistiche che usano le loro potenzialità, per migliorare la comunicazione. Il costruttore testo applica le sue competenze e migliora i livelli qualitativi del testo da redigere. Egli compone un testo secondo le tecniche delle grammatiche e cioè:

(...) grammatiche locali sono applicate ai testi per recuperare informazioni lessicali, morfologiche e sintattiche. realizzare l'analisi automatica di combinazioni di parole, ovvero l'analisi automatica di interi sintagmi e frasi complete. Non solo lessici elettronici, ma costruttori testi elettronici guidati dall'uomo, il linguista non si accontenta e cerca sempre più di realizzare, in modalità generazione, per produrre automaticamente parafrasi di frasi. La ricerca da me condotta parte dall'utilizzo delle frasi fisse, o ancor più dalla validità delle polirematiche per costruire un testo in parafrasi di frasi e li trasmette in tempo reale⁵.

4. Work Tools

Ultimo prodotto Work-Tools (2016)⁶ (attualmente in fase di sperimentazione) o *Linguistic Resources for Automatic Natural Language*. Esso è una proposta per una didattica inclusiva. L'utente si serve dei modelli interattivi a dominanza lineare, sui modelli ITS (*Intelligent Tutoring Sistem*), in cui l'uomo diventa costruttore

5 Cfr. per approfondimenti <http://www.nooj-association.org/media/k2/attachments/events/LiRANLG.htm>

6 *Work-Tools* © 2016 R. Bucciarelli; P. Villari; F. Terrone; M. Terrone; F. Santoro; R. Marcone. All. Rights reserved

delle proprie conoscenze. È stato concepito sul modello *Acro-Word*, ma contiene all'interno del *DB* anche oggetti in cui è possibile inserire video, filmati sequenze di frasi utili per la trascrizione e per la riconversione per translitterazione o per formulazione. Riconverte il logografo in caratteri cinesi (in uso comune definiti "ideogrammi")⁷ (vedi Fig. 3).



Fig. 3. Work-Tools – Riconversione codice cinese

Conclusioni

I *software* sono le risultanze positive di ipotesi operative condotte sul campo nel corso della ultradecennale esperienza di anni di studio e di lavoro di un gruppo di volontari e sono:

- *Type-Race* (2000), che è il primo prodotto a scopo didattico per supportare i discenti disabili sensoriali nelle consultazioni delle polirematiche e delle frasi fisse e di ampliare la conoscenza dei linguaggi professionali.
- *Acro-Word* (2013), che è il primo vero costruttore testo⁸ composto da:
 1. un corpus linguistico con un *DB* di oltre 5000 tra frasi fisse e frasi fatte rilevate dal dizionario di consultazione di Tullio De Mauro (1999);
 2. un corpus di scrittura (foglio di scrittura), in cui interviene l'uomo e produce, riformula e trasforma un testo avvalendosi della tecnica del richiamo e delle sue competenze linguistiche;
 3. un corpus di trascrizione e di trasformazione e di riformulazione del testo scritto dall'uomo in fase 2. Queste due fasi si compensano ed è un prodigio umano, perché l'acronimo si trasforma e traduce in frase che va a comporre un testo gestito dall'uomo. Inoltre, è in questa fase che rivediamo l'esatta riformulazione della frase per supportare la traduzione, ma soprattutto per costruire testi professionali in parafrasi.
- *Work-Tools* (2016), che si differenzia dall'*Acro-Word* per la sola composizione del *DB*, perché c'è l'inserimento dell'oggetto che ti permette di inserire sequenze filmate utile per la riconversione dei linguaggi segnici e per le lingue con codici composti da ideogrammi (codice cinese).

⁷ *Ideogramma*: i linguisti tendono ad escludere l'opinione che un sistema di scrittura possa codificare direttamente le idee senza avere da tramite la lingua e, quindi, generalmente preferiscono scartare i termini "ideogramma" e "pittogramma", nonostante siano spesso comunemente usati.

⁸ Ispirato agli studi di Silberstein, M. (2004).

Riferimenti bibliografici

- Arcuri, A., & D'Agostino, M. (1982). La forma haiu a+ infinito nel sistema verbale del siciliano. In *Linguistica contrastiva. Atti del XIII Congresso Internazionale di Studi della Società di Linguistica Italiana (Asti, 26-28 Maggio 1979)* (pp. 359-369). Roma: Bulzoni.
- Bucciarelli, R. (2015). La lessicografia per la descrizione e l'analisi dei corpus linguistici. *Formazione & Insegnamento. Rivista internazionale di Scienze dell'educazione e della formazione*, 12(4), 389-396.
- Bucciarelli, R. (2016). *I sentieri del cuore*. Salerno: I.R.I.S.
- D'Agostino, E., & Elia, A. (1998). Il significato delle frasi: un continuum dalle frasi semplici alle forme polirematiche. In AA. VV, *Ai limiti del linguaggio* (pp. 287-310). Bari: Laterza.
- De Bueris, G., & Elia, A. (2008). *Lessici elettronici e descrizioni lessicali, sintattiche, morfologiche ed ortografiche*. Salerno: Plectica.
- De Mauro, T. (1999). *Grande dizionario italiano dell'uso*. Torino: UTET.
- Elia, A., D'Agostino, E. & Martinelli, M. (1985). Tre componenti della sintassi italiana: frasi semplici, frasi a verbo supporto e frasi idiomatiche, in *Sintassi e morfologia della lingua italiana d'uso. Teorie e applicazioni descrittive*. In *Atti del XVII congresso internazionale della Società di Linguistica Italiana (Urbino, 11-13 settembre 1983)* (pp. 311-325) (a cura di A. Franchi De Bellis & L.M. Savoia). Bulzoni: Roma.
- Gross, M. (1988). Sur les phrases figées complexes du français. *Langue française*, (77), 47-70.
- Gross, M. (1989). La construction de dictionnaires électroniques. *Annales des télécommunications*, 44, 1-2, pp. 4-19. Springer: Verlag.
- Gross, M. (1990). La caractérisation des adverbes dans un lexique-grammaire. *Langue française*, (86), 90-102.
- Gross, M. (1995). On counting meaningful units in texts. *JADT*, 5-18.
- Harris, Z. S. (1970). From phoneme to morpheme. *Papers in Structural and Transformational Linguistics* (pp. 32-67). Springer: Dordrecht.
- Longobardi, G. (2011). A minimalist program for parametric linguistics. *Organizing Grammar: Linguistic Studies in Honor of Henk van Riemsdijk; Broekhuis. H*, 407-414.
- Silberztein, M. (2004). NooJ: A Cooperative, Object-Oriented Architecture for NLP. *INTEX pour la Linguistique et le traitement automatique des langues*. Cahiers de la MSH Ledoux, Presses Universitaires de Franche-Comté.

