

Annotazione e Analisi Sintattica del parlato afasico

di Carmela Sammarco*

Abstract

Syntactic analysis and annotation of aphasic speech

The present work consists of the analysis and syntactic annotation (parsing) of a corpus of spoken Italian consisting of *Cookie Theft Picture* descriptions produced by three aphasic speakers and three control subjects. The corpus consists of 610 words corresponding to 8 minutes and 55 seconds of recorded speech, transcribed and manually annotated with ANANAS L2 (*Annotazione Analisi Sintattica*, for learners of Italian as foreign language).

The annotation allowed to 1) compare the corpus with other data belonging to more than one linguistic variety and produced by speakers with different communicative abilities (e.g. learners of Italian as a foreign language); 2) to take into account phenomena of spoken without considering them as deviant structures, 3) to annotate any impaired structures, 4) to indicate the syntactic constituent in which the impaired structure is located.

Impaired structures were labeled on the basis of two dimensions: the type of impairment (omission, substitution, extra element) and the syntactic constituent and grammatical category in which they recur. The labeling of the corpus revealed that deviations pertain to different levels of syntactic construction and that a deviation pertinent to one syntactic level is not an obstacle to the non-impaired construction of the higher syntactic unit.

Keywords: Aphasic speech, Syntactic analysis, Annotation of impaired structures.

I Introduzione

Lo scopo del seguente lavoro è mostrare come l'applicazione del sistema AN.ANA.S L2 (Annotazione e Analisi Sintattica per apprendenti di italiano come L2) (Turco, Voghera, 2010) al parlato patologico permetta di avere dati quantitativi e qualitativi sulle caratteristiche sintattiche sul parlato afasico, sul tipo di errore e sul dominio sintattico in cui esso occorre. In questo lavoro si presenta l'analisi e l'annotazione sintattica di un corpus di italiano parlato semispontaneo di tre pazienti afasici e di tre informanti che costituiscono un gruppo di controllo. Il corpus presenta un tipo di elocuio elicito e consiste nelle descrizioni orali della vignetta nota come *Cookie*

* Docente a contratto di Linguistica Generale e Didattica delle Lingue Straniere, Dipartimento di Studi Umanistici, Università degli Studi di Salerno.; csammarco@unisa.it.

Theft Picture (Goodglass, Kaplan, 1983) utilizzata nel test BADA (Miceli *et al.*, 1994) per testare la fluenza e la capacità di produzione orale da parte dei pazienti afasici¹. Le descrizioni dei pazienti e dei soggetti di controllo sono state dapprima trascritte ortograficamente secondo le norme CLIPS (Savy, 2007) e in seguito annotate manualmente attraverso il sistema di annotazione sintattico AN.ANA.S L2, sviluppato dall'osservatorio linguistico nazionale Parlaritaliano² dell'Università di Salerno.

Ideato per annotare le strutture dell'interlingua (Selinker, 1972) in testi sia orali sia scritti, il sistema di annotazione AN.ANA.S L2 è stato progettato per etichettare le strutture che si allontanavano dalla lingua di arrivo (*pre-target*). Essendo stato utilizzato per l'annotazione di strutture non grammaticali, è stato applicato lo stesso sistema di annotazione, apportando modifiche nella DTD (*Document Type Definition*) laddove necessario, per annotare deficit linguistici o strutture deficitarie (SDEF) di parlanti afasici, ovvero strutture riconoscibili come scelte linguistiche poco probabili, o del tutto improbabili, indipendentemente dal contesto e dalla modalità comunicativa, e che sono la manifestazione di un sistema linguistico danneggiato. Inoltre, come le strutture *pre-target* delle produzioni degli apprendenti, anche quelle dei parlanti afasici possono rivelare aspetti importanti dei meccanismi comunicativi generali (Nuzzo, Favilla, 2015; Favilla, Ferroni, 2018).

A tal proposito, lo schema di annotazione utilizzato in AN.ANA.S L2 si basa sul sistema AN.ANA.S (Cutugno, Voghera, 2004) che si basa su categorie di linguistica generale e offre la possibilità di cambiare il *tagset* di definizione anche durante l'annotazione a seconda delle diverse analisi richieste. In questo modo permette di avere dati comparabili e verificabili di varietà di lingua prodotte da parlanti con competenze comunicative di vario tipo.

Le categorie descrittive alla base di AN.ANA.S permettono di tener conto anche di fenomeni frequenti nei testi parlati (Voghera *et al.*, 2005; Voghera, 2017; Bazzanella, 1992). Ormai è riconosciuta l'importanza del parlato nella ricerca sull'afasia (Armstrong, 2000; Linnik *et al.*, 2016), non solo per il fatto che esso rappresenta il dominio di indagine più naturale dal punto di vista filogenetico e ontogenetico (De Mauro, 2008; Voghera, 2017), ma anche perché offre una migliore individuazione delle SDEF. L'analisi del parlato, infatti, permette di distinguere le SDEF da strutture come pause, disfluenze, ripetizioni, cambiamenti di programmazione sintattica che sono frequenti nel parlato ma non sono da considerarsi strutture deficitarie. Nella modalità parlata, infatti, la quasi contemporaneità dei tempi di programmazione e produzione porta i parlanti a scegliere strutture linguistiche funzionali a farsi comprendere dal proprio interlocutore nel minor tempo possibile. Di conseguenza, tali fenomeni sono strategie utili a chi costruisce il testo *in fieri*. Il ricorso a tali scelte può essere più o meno frequente a seconda della situazione enunciativa, dello scopo del testo e delle caratteristiche individuali dei parlanti (Voghera, 2017).

Sebbene sia riconosciuta la primarietà del parlato nella ricerca sui disturbi del linguaggio sono pochi i corpora in diverse lingue consultabili che offrono annotazioni

di fenomeni riguardanti diversi livelli linguistici e un possibile confronto interlinguistico dei risultati (Kasselimis *et al.*, 2020; Thompson *et al.*, 2003; Edwards, 2005). Per l’italiano ricordiamo, oltre al progetto di *Aphasia Bank*³ annotato per parti del discorso (POS), il corpus *Aphasia*⁴ dell’Università di Bologna che raccoglie trascrizioni di parlato patologico di diversa natura, e il più recente corpus *CIPPS* (Dovetto *et al.*, 2017) che raccoglie produzioni di parlanti schizofrenici trascritte e annotate su diversi livelli linguistici. Infine, ricordiamo anche il corpus del progetto *Oplon* di Calzà *et al.* (2021) raccolto per la creazione di un sistema di annotazione automatica per il riconoscimento dei disturbi cognitivi di media gravità. Per quanto riguarda le altre lingue segnaliamo corpora di parlato afasico di parlanti olandesi (Westerhout, Monachesi, 2006), di cinese cantonese (Kong, Law, 2019), di russo (Khudyakova *et al.*, 2016), di croato (Kuvač Kraveljević *et al.*, 2017), di greco (Varlokosta *et al.*, 2016) e di inglese (Mirman *et al.*, 2010; Williams *et al.*, 2010; MacWhinney *et al.*, 2012; Laures-Gore *et al.* 2016).

Il contributo che AN.ANA.S può dare alla ricerca sul parlato afasico è quello di offrire un’annotazione *parsing* che non solo assegna una categoria alle diverse SDEF, operazione tra l’altro molto problematica (Andorno *et al.*, 2017; Kasselimis *et al.*, 2020), ma individua anche il dominio sintattico in cui esse occorrono. Inoltre lo schema di annotazione utilizzato in AN.ANA.S L2, che si basa sul sistema AN.ANA.S, utilizza categorie che si rifanno alla linguistica funzionale (Halliday, 1985) che permettono di descrivere in maniera oggettiva i fatti linguistici analizzati e offre la possibilità di cambiare il *tagset* di definizione anche durante l’annotazione a seconda delle diverse esigenze delle analisi. In questo modo permette di avere dati comparabili e verificabili di varietà di lingua prodotte da parlanti con competenze comunicative di vario tipo. I parametri utilizzati dal sistema sono riconosciuti da iniziative internazionali come la Text Encoding Initiative (TEI) e attraverso il software *XGate* che permette di generare file con un linguaggio standard XML (*eXtensible Mark-up Language*) dando la possibilità di visualizzare l’organizzazione gerarchica dei testi annotati e di fare *queries* mirate per avere dati quantitativi e qualitativi (Cutugno, Voghera, 2004; Voghera *et al.*, 2004; 2005).

2 Il parlato afasico

Le afasie di Wernicke e di Broca rappresentano due macrocategorie a cui si riconducono le numerose e diverse patologie cognitive che compromettono il linguaggio e che dal punto vista eziologico sono associate a lesioni acquisite in diverse aree cerebrali dell’emisfero sinistro. In linea generale, nel caso dell’afasia di Broca si tratta essenzialmente di lesioni al lobo frontale, nel terzo giro della corteccia cerebrale o zona prerolandica; nella afasia di Wernicke, invece, generalmente le lesioni riguardano l’area del lobo temporale sinistro, vicino alle aree della corteccia occipitale.

Sebbene, grazie all'incremento di studi sul parlato afasico, sia ormai riconosciuto che i pazienti appartenenti a diverse categorie diagnostiche spesso presentano molti deficit linguistici simili (Basile, 2014) e che non c'è una corrispondenza biunivoca tra lesione e deficit linguistico (Kassellimis *et al.*, 2017), in linea generale i pazienti con afasia di Broca presentano: una riduzione della lunghezza media delle frasi (*Mean Length of Utterance* o *MLU*), la compromissione dei tratti prosodici, un maggiore uso di parole appartenenti a categorie lessicali aperte (es. nomi, verbi, aggettivi e avverbi), l'uso di verbi al modo infinito, l'omissione e la sostituzione di morfemi grammaticali (Denes, 2009). I pazienti con afasia di Wernicke, invece, si distinguono soprattutto per avere un eloquio fluente (Edwards, 2005) e un'eccessiva verbosità, in quanto, a differenza degli afasici di Broca, non hanno compromesse le abilità articolatorie del linguaggio⁵. Inoltre, in aggiunta alle difficoltà nell'uso della morfologia flessiva e nella costruzione di testi sintatticamente complessi, questi pazienti mostrano problematicità nel recupero di parole appartenenti a classi lessicali aperte (Edwards, 2005). I pazienti con afasia di Wernicke, infatti, manifestano serie difficoltà nel denominare persone ed eventi, ciò li porta ad avere forti esitazioni nel momento in cui devono utilizzare dei nomi. In compenso usano appropriatamente le parole funzionali e ordinano in maniera corretta le parole nei testi che producono. Dal punto di vista ricettivo, hanno la comprensione compromessa, in quanto non riconoscono più le immagini uditive delle parole, ovvero hanno difficoltà a dare significato alle sequenze foniche che ascoltano (Basile, 2012a, p. 37).

3

L'analisi del parlato e la considerazione delle strutture deficitarie

L'analisi di un corpus di parlato afasico richiede necessariamente di distinguere le caratteristiche che costituiscono un normale testo parlato da quelle che possono essere classificate come SDEF (cfr. PAR. 1).

Spesso il parlato è stato descritto come varietà linguistica che presenta strutture frammentarie, disordinate e instabili, definite "non canoniche" in relazione ad una norma di riferimento che si rifa soprattutto al modello della lingua scritta. Tuttavia è ormai risaputo che il parlato si caratterizza per la frequenza di fenomeni e scelte linguistiche che sono il prodotto dell'intersezione di fattori che costituiscono la modalità parlata ovvero: 1. il canale fonico-uditivo e visivo-gestuale, 2. la compresenza del parlante e del ricevente e 3. la simultaneità dei tempi di programmazione e ricezione del messaggio (Voghera, 2017). L'interazione di questi tre fattori determina scelte linguistiche che i parlanti fanno perché più ottimali per la gestione dei tempi brevi di programmazione e di ricezione e per il raggiungimento dei propri scopi comunicativi.

I testi parlati presentano quindi peculiarità su tutti i livelli linguistici. In modo molto semplificato riportiamo nella tabella 1 i principali fenomeni che caratterizzano i testi parlati per ciascun livello linguistico:

TABELLA I

Fenomeni che maggiormente caratterizzano i testi parlati a livello sintattico e testuale

1) Livello testuale	Uso della prosodia per connettere porzioni di testo, ripetizioni, disfluenze, segnali discorsivi, ridondanza, deissi.
2) Livello frasale	Tendenzialmente frasi e sintagmi brevi nei testi dialogici e più lunghi nei testi monologici, presenza di frasi senza verbo, espressione delle relazioni di subordinazione attraverso marche prosodiche.
3) Livello lessicale	Scelte di parole polisemiche, vaghe.

Gli elementi riportati nella TAB. I descrivono le scelte funzionali che i parlanti fanno nella modalità parlata in cui utilizzano il canale fonico-uditivo e visivo-gestuale, interagiscono in maniera immediata con il ricevente e hanno tempi di programmazione e produzione ridotti.

Oltre alle caratteristiche legate ai fattori che costituiscono la modalità parlata, vanno tenute presenti le condizioni legate alla situazione enunciativa e alle caratteristiche e competenze dei singoli parlanti che possono determinare la maggiore o minore frequenza di alcuni dei fenomeni visti nella TAB. I. In altre parole, data una situazione enunciativa le strutture scelte da un parlante hanno maggiori o minori probabilità di presentare determinate caratteristiche sulla base della maggiore o minore possibilità di programmare il testo, di interagire e co-costruire il testo con l'interlocutore.

Ad esempio, la frequenza di disfluenze considerata isolatamente non è indicativa della maggiore o minore capacità comunicativa piuttosto di una diversa capacità di gestire la programmazione del testo, la quale va messa in relazione a fattori legati alla maggiore o minore libertà di partecipazione alla conversazione e a fattori individuali, come l'età. Proprio a proposito dell'età, a titolo di esempio si veda il confronto tra le due narrazioni di una stessa storia prodotte da due parlanti non afasiche di età diverse: otto e ventisei anni.

Le due parlanti in (1) narrano la scena del rapimento del bambino del film *Il monello* di Charlie Chaplin: in entrambe le narrazioni troviamo fenomeni di disfluenza come 1. le pause (evidenziate in grassetto), 2. interruzioni (segnalate con il simbolo +), 3. cambiamenti di progettazione sintattica (in maiuscolo), 4. esitazioni, allungamenti vocalici o anche pause piene (in corsivo), 5. ripetizioni (*si tre o quattro signori, di tre signori; questi signori, che litigava con con gli altri signori// < //che litigava con gli altri signori* nella parlante di otto anni; *questo bambino venga sottratto// diciamo<oo> rapito// in qualche modo sottratto// a charlot//* nella narrazione della parlante universitaria, quella di ventisei anni). Tuttavia ciò che cambia è la frequenza con cui tali fenomeni occorrono nei due testi. Nella narrazione della bambina me-

diametralmente ci sono delle pause ogni sette parole circa, mentre nella narrazione della parlante adulta ogni tredici. Inoltre abbiamo una sola pausa lunga che occorre nella narrazione della bambina dopo la sequenza più lunga di tutto il suo eloquio, come se appunto la programmazione della frase avesse richiesto più tempo per andare avanti. La maggiore frequenza dei fenomeni di disfluenza nella parlante più piccola è indicativa di una maggiore necessità di prendere tempo per riprogrammare il discorso.

(1)

Bambina 8 anni

ho visto questo video in bianco e nero// <pb>
 non parlavano //<pb>// SO-
 PRATTUT+ POI// <pb> //eb
 raccontava di tre signori di tre o
 quattro signori che appunto//
 <pb> //che litigavano fra di
 loro //e un bambino che pian-
 geva //<pb> //QUESTI SIGNORI
 <pb> eh <pb> //tra uno di questi
 signori uno era il padre//<pb>
 //il bambino piangeva perché ap-
 punto vedeva il padre che litigava
 con con gli altri signori// <pl> //
 che litigava con gli altri signori
 <pb> //

Studentessa universitaria⁶

allora// <pb> quindi il film che ho visto// è un film
 tratto da<aa> //<ehm> da charlie chaplin// intitolato
 il monello del millenovecentoventuno //<pb> ED
 È<èè> <ehm> // diciamo s'intitola appunto il mo-
 nello sottratto a charlot //<pb> E PARLA DI QUESTO
 BAMB+// cioè almeno la<aa> la scena iniziale//<pb>
 con cui si apre il film //<ehm> mostra chiaramente
 CHE <ehm> / COME questo bambino venga sottratto
 diciamo<oo>// rapito //in qualche modo sottratto a
 charlot //<pb>// e <ehm>// diciamo che la scena<aa>
 / perlomeno la<aa> la scena iniziale del film è ambien-
 tata in una stanza//<pb>// in una camera// <pb> //
 e all'improvviso fanno irruzione in questa camera pro-
 prio dei<ii>// insomma non so//delle persone non so
 chi siano// però//anche poi la polizia che interviene
 all'improvviso//

Nel caso dei parlanti afasici ad essere rilevante non è solo la frequenza dei fenomeni di disfluenza, ma è anche la mancata riprogrammazione del testo. Nel parlato patologico, infatti, la gestione di questi meccanismi di riparazione e di programmazione spesso è senza fini progettuali (Basile, 2017). A tal proposito riportiamo l'esempio (2) tratto dal nostro corpus.

(2)

P: l'acqua viene fuori di più viene fuori tutt+ {[NOISE] può andare su_i piedi <eeh> st+ st+
 lei sta <sp> <mmm> e ri+ <mmm>
 E: cosa sta facendo la<aa>}
 P: la mamma sta asciugando il lavan+ sta asciugando il lavand+ <sp> sta asciugando <sp> eh
 la+ sta asciugando i piatti poi<ii> <lp>

In (2) la paziente ha difficoltà ad andare avanti e produce diverse pause piene (in grassetto e tra parentesi uncinate) e parole frammentate. Come vediamo dall'esempio, le disfluenze in questo caso non portano a una riprogrammazione del testo. Infatti, la parlante riesce a continuare il discorso solo in seguito alla sollecitazione dell'esaminatrice.

Tali premesse hanno lo scopo, quindi, di fornire un punto di riferimento più oggettivo per l'analisi delle SDEF nel caso del parlato patologico. Usando come metro di paragone il parlato semispontaneo, molte strutture considerate marginali nella teoria linguistica (come frasi senza verbo, cambi di progettazione sintattica) non sono valutate come anomalie, ma come appartenenti a fenomeni frequenti nel parlato non afasico (esempi 3 e 4 tratti dal corpus dei soggetti di controllo):

(3)

Una famiglia praticamente <lp> in cucina.

(4)

Che<ee> la mamma o<oo> non so chi è la mamma mhmh dice stai attento a non cadere <lp> no?

Gli esempi (3) e (4) riportano estratti delle narrazioni fatte dai soggetti di controllo. In (3) troviamo una frase senza verbo e una pausa lunga, in (4) troviamo l'uso del connettivo *che* con funzione quotativa che introduce una nuova informazione ma non connette la frase a strutture precedenti.

Concludendo, i fenomeni come i cambi di progettazione sintattica con riparazioni testuali (Basile, 2014; 2015), le ripetizioni, le esitazioni sono stati annotati utilizzando le etichette già presenti per l'annotazione del parlato (Voghera, Cutugno, 2004) (cfr. PAR. 4) e non sono stati segnalati come SDEF. Allo stesso modo non sono stati considerati come SDEF e non hanno ricevuto un trattamento a parte le scelte linguistiche riconducibili a varietà diatopiche (Dovetto *et al.*, 2017).

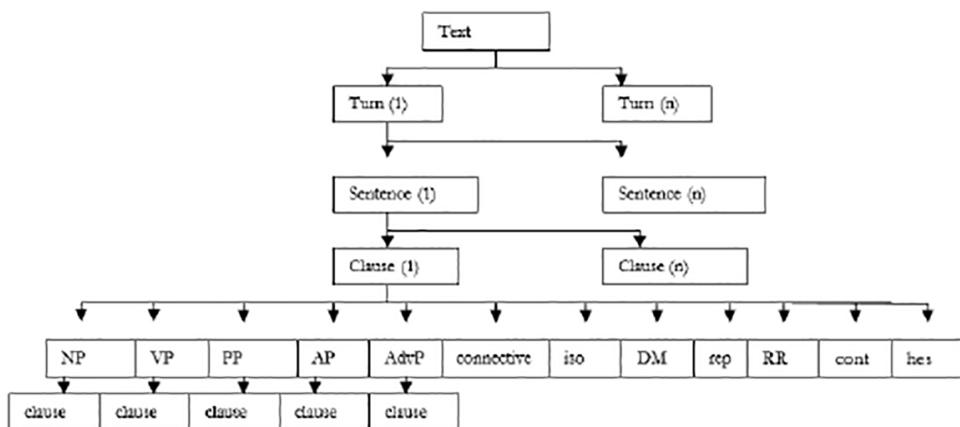
4

Il sistema di annotazione

AN.ANA.S è nato come un sistema *treebank* per annotare sintatticamente produzioni parlate e scritte (Turco, Voghera, 2010). L'annotazione e l'analisi attraverso AN.ANA.S permette di creare testi in cui le relazioni gerarchiche tra i sintagmi e i tipi di livelli sintattici siano visibili e siano rese anche graficamente attraverso alberi sintattici. I testi risultano organizzati in unità linguistiche gerarchicamente ordinate secondo i diversi livelli sintattici: 1. testo, 2. turno o paragrafo, 3. frase, 4. clausola, 5.

sintagma nominale (NP), sintagma verbale (VP), sintagma preposizionale (PP), Sintagma avverbiale (AdvP), segnali discorsivi (DM), ripetizioni (REP), cambiamenti sintattici (RR), sintagmi discontinui (CONT), esitazioni (HES). Lo schema di annotazione è rappresentato nella FIG. I.

FIGURA I
Rappresentazione dello schema di annotazione in AN.ANA.S.



Ogni entità (testo, turno, frase, clausola ecc.) rappresenta un nodo, per ognuna di esse sono previsti una serie di attributi attraverso i quali è possibile indicare i valori per i parametri di ciascuna unità. Ad esempio, per ogni sintagma si può indicare la relazione grammaticale che ha all'interno della frase e le caratteristiche lessicali legate alla testa (ad esempio, se è un nome polirematico) oppure per ogni clausola è possibile indicare se è una principale, una subordinata o una senza verbo.

L'albero sintattico creato attraverso l'annotazione è in formato XML (Cutugno, Voghera, 2004) e ogni entità è associata a un *tag* secondo gli standard XML⁷. Il risultato finale è un testo annotato nei vari livelli sintattici e graficamente strutturato secondo le relazioni gerarchiche di ciascun livello. Nella FIG. 2 riportiamo un estratto dell'annotazione di uno dei testi che formano il corpus.

FIGURA 2
Rappresentazione di un turno annotato con AN.ANA.S L2

```

<turn_compl="1" turn_id="p2">
  <sentenza n_of_sentences="2">
    -<clause type="m" n_of_phrases="4">
      <CVB mod="I" discontinuous="F" position="post" mrv="I" sub="I" dia="a" n_of_arg="1" cop_vb="I" lexeme="succedere" sub_type="a"> succede </CVB>
      <NP mod="I" mod="I" det="I" discontinuous="F" position="post" pred="I" n="I" mrv="I" sub="I" lexeme="guado" obj="I"> un grosso guado </NP>
      <NP mod="I" mod="I" det="I" discontinuous="F" position="post" mrv="NP" pred="I" n="I" mrv="I" complement="mamma" head="con" mult_n="2" pre="dev" wf="I">
        con la mamma
      -<clause type="dep" n_of_phrases="1" arg="I" link="rel">
        <NP multiple="I" mod="I" det="I" discontinuous="F" position="pred" pred="I" n="I" mrv="I" sub="I" lexeme="che" obj="I"> che </NP>
      -<clause type="dep" n_of_phrases="2" arg="I" link="compl">
        <CONNECTIVE type="verb" p_of_prep="preposizione"> mentre </CONNECTIVE>
        <VP mod="I" discontinuous="F" mrv="I" sub="I" dia="a" n_of_arg="2" cop_vb="I" lexeme="lavare">
          sta lavando
        -<NP mod="I" mod="I" det="I" discontinuous="F" position="post" pred="I" n="I" mrv="I" sub="I" lexeme="piatto" obj="I"> i piatti </NP>
        </VP>
      </NP>
    </clause>
  </sentenza>
</turn>

```

Nella FIG. 2 è riportata l'annotazione dell'inizio del turno di un paziente *succede <uu> un grosso guado con la mamma che} mentre sta lavando i piatti <eh> dimentica di chiudere il rubinetto per l'acqua*. Come si evince dalla figura, il testo annotato mostra che il turno è il livello testuale che contiene le altre unità, ovvero la frase (*sentence*), la clausola (*clause*), e i vari sintagmi che a loro volta possono essere sovraordinati ad altre unità come la clausola, come *che mentre sta lavando i piatti dimentica di chiudere il rubinetto per l'acqua*.

Le etichette utilizzate in AN.ANA.S L2 per l'annotazione sintattica sono forgiate sulle categorie sintattiche della linguistica funzionale (Halliday, 1985) e sono finalizzate ad una descrizione generale in modo da risultare adeguate all'annotazione di diversi tipi di testi e di SDEF. Inoltre, lo schema di annotazione prevede l'etichettatura di strutture non canoniche (“wf=f”). Nella FIG. 2, ad esempio, è possibile visualizzare l'annotazione della SDEF *il rubinetto per l'acqua*, come sostituzione (*Dev* [cfr. PAR. 5]) della preposizione *di*.

5 Tipi di SDEF e livelli sintattici

Nel corpus analizzato le occorrenze di SDEF sono riconducibili a tre macrocategorie (Turco, Voghera, 2010): 1. l'omissione di elementi (*Miss*) ovvero, quando i parlanti non producono un elemento che ci si aspetterebbe nella sequenza target, come nell'esempio (5); 2. l'aggiunta di elementi (*Extra*), come in (6); 3. la sostituzione di una struttura target con un elemento appartenente alla stessa categoria grammaticale o anomalie nella costruzione sintattica (*Deviation, Deviation syntactic structures*) (cfr. [7]).

(5)

l'acqua dal piano *va distribuirsi*
su tutto il pavimento'. invece di (?) l'acqua dal piano *va a distribuirsi*
su tutto il pavimento'

(6)

e la bambina ride <sp> che
suo il fratello è caduto <lp>. invece di (?) e la bambina ride <sp> che (*suo*)
(*il*) fratello è caduto <lp>.

(7)

sta succedendo un grosso guaio
con la mamma dimentica di chiudere il rubinetto *per* l'acqua. invece di (?) sta succedendo un grosso guaio
alla mamma dimentica di chiudere il rubinetto *dell'*acqua.

In (5) la sequenza *va distribuirsi* è stata annotata come omissione, in quanto struttura mancante della preposizione *a* retta dal verbo *andare*, in (6) è stato annotato con il tag *Extra*, il SN *suo il fratello* in quanto troviamo un doppio determinate laddove sarebbe molto più probabile trovarne solo uno trattandosi di un SN che ha come testa un nome parentale⁸; infine, in (7) sono state annotate come SDEF devianti le preposizioni *con* e *per* laddove ci aspetteremmo le forme articolate *alla* e *del*.

TABELLA 2
Annotazione di SDEF per livello sintattico

Livelli sintattici	Tipi di SDEF
Testuale	Non pertinenza del discorso con il compito richiesto
Frase	Parafasie che comprendono più clausole Omissione/Sostituzione/Extra/Congiunzione tra clausole Omissione/Sostituzione/Extra/Coreferenza
Clausola	Devianza nell'ordine dei sintagmi Omissione/Extra marche di accordo tra sintagmi Omissione/Sostituzione/Extra congiunzione o preposizione tra sintagmi Parafasie che comprendono più sintagmi
Sintagma	Omissione/Sostituzione/Extra determinante Omissione/Sostituzione/Extra modificatore Omissione/Sostituzione lessema Omissione/Sostituzione/Extra preposizione Omissione/Sostituzione morfema di accordo

Oltre al tipo di SDEF il sistema di annotazione AN.ANA.S L2 permette di segnalare come le SDEF riguardano sia i rapporti tra clausole (livello frasale) sia le relazioni degli elementi all'interno delle clausole (livello clausale) sia le relazioni degli elementi all'interno dei singoli sintagmi (livello sintagmatico e lessicale). Inoltre, è possibile annotare SDEF a livello testuale, laddove ci siano anomalie che riguardano la pertinenza dell'argomento al compito richiesto, e l'uso anomalo di connettivi che legano diverse frasi. Nella TAB. 2 sono illustrati i tipi di SDEF per livelli sintattici.

Sono stati segnalati come SDEF al livello di frase i casi di parafasie, come quella nell'esempio (8), o di omissione/sostituzione/aggiunta di elementi tra le clausole.

(8)

sì cioè io ho cercato perché fosse una cosa da fare non molto veloce in un certo senso che non fosse<ee> da leggere i<ii> romanzi per

La sequenza in (8) sembra contraddittoria: non si capisce se il paziente pensasse che fosse un compito non troppo veloce leggere romanzi o il contrario. La presenza di due negazioni porta a considerare l'intera struttura come un caso di SDEF con deviazione sintattica.

Allo stesso modo eventuali anomalie tra i sintagmi, hanno costituito motivo di considerare il livello *clause* non correttamente formato (cfr. [9]).

(9)

e così il bambino prende i
biscotti dal <sp> dal <ss> se
<sp> se raccolt+ raccolt+ dal
vasetto.

invece di (?)

e così il bambino prende i
biscotti dal <sp> dal <ss> se
<sp> se raccolti raccolti dal
vasetto.

In (9) la sequenza presenta un caso di mancanza di accordo tra la forma participiale del verbo *raccogliere* e il SN *i biscotti*. Questo tipo di SDEF è stata annotata come omissione di accordo nei valori che riguardano l'annotazione del SV e allo stesso tempo anche la clausola in cui occorre è stata annotata con lo stesso valore.

Infine, all'interno dei sintagmi sono state considerate le SDEF relative alle occorrenze extra o alle omissioni di determinanti e modificatori. Ad esempio riproponiamo l'esempio (6) in (10).

(10)

e la bambina ride <sp> che
suo il fratello è caduto <lp>.

invece di (?)

e la bambina ride <sp> che (*suo*)
(*il*) fratello è caduto <lp>.

L'elemento 'Extra' che, come abbiamo già detto, è rappresentato dall'occorrenza di due determinanti *il* e *suo* costituisce una SDEF di livello sintagmatico.

Come è risaputo, non è semplice stabilire la natura della SDEF e non sempre è inequivocabile il livello sintattico e le categorie a cui essa afferisce (Andorno *et al.*, 2017; Kasselimis *et al.*, 2020). Come abbiamo già visto, commentando l'esempio (8), in una sequenza le SDEF possono essere relative a più forme target possibili. Cioè, a seconda della forma target che si ritiene più probabile, è possibile identificare una SDEF. Inoltre, una stessa SDEF coinvolge più livelli sintattici: come abbiamo visto anche a proposito di (9), le SDEF che riguardano la morfologia verbale coinvolgono il sintagma e la costruzione della clausola allo stesso tempo. Un ulteriore esempio è dato da (ii), dove troviamo una forma infinita laddove ci si aspetterebbe un verbo in forma finita e viceversa si produce una forma finita laddove invece potrebbe esserci una forma infinita se occorresse un verbo modale.

(ii)

la madre <sp> *capire* che chiude
de l'acqua e quindi

invece di (?)

la madre <sp> *capisce* che deve
chiudere l'acqua e quindi

In (ii) è possibile individuare almeno tre SDEF: l'omissione di accordo in *capire*, l'omissione di una probabile forma modale come *deve*, *può* ecc., una devianza nell'uso della forma finita *chiude*. Le tre forme di SDEF, a loro volta, rimandano a SDEF sul livello clausale, per la mancanza di accordo tra il probabile soggetto *la madre* e la forma verbale *capire*, ma anche al livello di frase, in quanto la forma finita *chiude* nella clausola subordinata può essere indicativa di una problematicità a costruire la coreferenza tra il soggetto della clausola oggettiva e della principale.

6 Il corpus

Il corpus analizzato consiste nelle descrizioni della vignetta nota come *Cookie Theft Picture* (Goodglass, Kaplan, 1983), usata in ambito clinico per l'elicitazione di eloquio semi-spontaneo, da parte di tre pazienti. Ai testi prodotti dai pazienti sono state affiancate tre descrizioni da parte di soggetti di controllo, al fine di rendere possibile una valutazione più oggettiva delle descrizioni prodotte.

Il corpus analizzato consta di 610 parole totali, che sono distribuite come riportato nella TAB. 3.

TABELLA 3
Totale del numero delle parole per i parlanti presenti

	Numero partecipanti	Numero parole	Tempo
Pazienti	3	397	6':50''
Soggetti di controllo	3	213	2':05''
TOT	6	610	8':55''

Ciascun paziente mediamente ha prodotto circa 132 parole, mentre in media si hanno 70 parole per ciascun parlante del gruppo di controllo. Anche la durata delle descrizioni dei pazienti è molto più lunga rispetto a quella dei soggetti di controllo. La descrizione di ciascun controllo non supera il minuto, mentre quella di ciascun paziente mediamente dura circa due minuti.

Nella TAB. 4 si riportano le informazioni dei pazienti e dei membri dei gruppi di controllo, riguardanti, il sesso, l'età, gli anni di scolarizzazione e l'eventuale dialettofonìa.

TABELLA 4
Dati relativi al sesso, all'età, alla scolarizzazione e alla provenienza dei pazienti e dei soggetti di controllo

Pazienti	Sesso	Età	Scolarità	Luogo di provenienza	Dialettofoni
V.C.	M	23	8	Campalto (VE)	sì
G.B.	M	66	13	Padova	sì
L.B.	F	55	8	Venezia	sì
Soggetti di Controllo					
T.B.	M	35	13	Pellestrina (VE)	sì
M.S.	M	42	13	Venezia	sì
I.B.	F	54	8	Venezia	sì

I pazienti hanno un'età compresa tra i ventidue e i cinquantacinque anni, due pazienti hanno un livello di istruzione di scuola secondaria di primo grado, soltanto uno presenta una scolarizzazione fino alla scuola secondaria di secondo grado. I pazienti selezionati per il nostro lavoro presentano buoni punteggi alla valutazione del linguaggio spontaneo, valutato attraverso il test di Aachen (AAT) utilizzato nella sua versione italiana (Luzzatti, Wilmer, Bleser, 1996). Tutti hanno registrato difficoltà nella comprensione scritta ma non in quella orale.

Il paziente V.C. ha subito un trauma cranico in seguito a un grave incidente risalente a cinque anni prima della raccolta dati. La valutazione neuropsicologica al momento del trauma era limitata solo al versante non verbale in quanto il paziente mostrava un grave deficit afasico. Al momento della somministrazione del test presenta secondo il test AAT, una situazione generale notevolmente migliorata, soltanto nella comprensione e nella scrittura ha ancora qualche problema. Nei test di parlato, il punteggio più alto è nella ripetizione. Nel parlato spontaneo presenta minimi segni di difficoltà articolatoria e l'eloquio rallentato. Il secondo paziente, G.B., presenta un danno da emorragia subaracnoidea in seguito a un aneurisma cerebrale dell'emisfero sinistro che risale all'anno prima dell'esperimento. In generale dimostra un livello medio di gravità nella comprensione orale e soprattutto nella comprensione scritta. I risultati del test AAT hanno rivelato una buona performance solo nella ripetizione, mentre nel test di denominazione dimostrano un disordine medio. Per quanto riguarda il linguaggio spontaneo, secondo i risultati del test AAT, non presenta difficoltà articolatorie, produce frasi lunghe e complesse con accavallamenti e ripetizioni di parti delle frasi, e interruzioni e errori nell'uso della morfologia flessiva. La terza paziente, L.B., al momento della ricerca presentava un trauma per un'emorragia supra-intraparenchimale in seguito a un aneurisma dell'emisfero sinistro, risalente all'anno prima della raccolta della raccolta dei dati. Dal test AAT emerge che l'abilità più compromessa è la comprensione scritta. La produzione orale nel *token test* e nella denominazione non presenta un punteggio molto alto. La ripetizione e la scrittura presentano invece un grado elevato di gravità. Il test sul parlato spontaneo mette in evidenza la tendenza a produrre strutture stereotipate e perseverazioni, tuttavia presenta comunque un punteggio alto. La conversazione spontanea risulta difficoltosa e richiede aiuto: emergono, infatti, alcune parafasie semantiche, produzioni *nonsense* e anomalie. L'eloquio presenta un ritmo lento con alcune difficoltà articolatorie.

7 Risultati

L'annotazione del corpus attraverso AN.ANA.S L2 ha permesso di considerare tre aspetti:

1. frequenza del tipo di struttura non target (omissione, sostituzione, elemento extra e inversione dell'ordine dei costituenti);
2. livello sintattico e categoria grammaticale in cui ricorre la SDEF;
3. la possibilità di avere dati quantitativi e qualitativi degli aspetti linguistici che caratterizzano il parlato afasico.

Per quanto riguarda la frequenza delle SDEF, il software XGate, attraverso cui è stata eseguita l'etichettatura manuale, ha permesso di interrogare il corpus e di costruire *queries* mirate. Nella TAB. 5 riportiamo il numero delle occorrenze delle SDEF e la loro percentuale calcolata sul totale.

TABELLA 5
Frequenza e tipo delle SDEF nel corpus dei parlanti afasici

Tipi di SDEF		%
Miss	II	44
Dev	I3	52
Extra	I	4
TOT	25	100

Le SDEF che sono più frequenti sono scelte lessicali poco probabili a cui seguono le omissioni di materiale e infine l'aggiunta di elementi linguistici che non sono classificabili come tentativi di riprogrammazione (ad esempio *il suo fratello*). Un altro tipo di *query* ci permette di conoscere quali sono le categorie grammaticali e i livelli sintattici in cui occorrono le SDEF. Nella TAB. 6 riportiamo la frequenza delle SDEF classificate sulla base della categoria grammaticale e lessicale.

Nella TAB. 6 è possibile notare innanzitutto che la maggior parte delle SDEF riguarda la categoria del nome e consiste nell'uso di sostantivi diversi da quelli che sono prevedibili in quel contesto, come si vedrà più in dettaglio nella TAB. 7: essi possono essere sia nomi che sono la testa di SN, sia nomi che costituiscono parte di SP. Inoltre è possibile anche notare che la maggior parte delle omissioni coinvolge la morfologia di accordo dei nomi e dei verbi: in quest'ultimo caso si manifesta nell'omissione delle marche di genere e numero delle forme participiali e nel ricorso di forme infinite al posto di forme coniugate (*capire* vs *capisce*). I casi di aggiunta di materiale (*extra*) riguardano soprattutto i determinanti.

TABELLA 6
Classificazione delle SDEF per tipo e categoria lessicale o grammaticale

	Det.	Nomi	Avv.	Verbi	Prep.	Morfologia di accordo	Tempo	Modo
Miss	I	I		I		7	I	
Dev		7	2		2		I	I
Extra	I							
TOT	2	8	2	I	2	7	2	I

La distribuzione delle SDEF per tipi di sintagmi e per livelli sintattici è riportata nella TAB. 7.

TABELLA 7
Distribuzione delle SDEF per tipi di sintagmi e livelli sintattici

Frase	Vp						Np			Pp			AdvP		
	Conj.	Lex	Aux	Tense	Mood	Accordo	Participio	Lex	Det	Accordo	Lex	Prep	Det	Lex	TOT
Miss		I	I			2		3		I	I				9
Dev	I			I	I				3			4	2	I	2
Extra									I						I
Tot	I		I	2	I	2		3	3	I	I	5	2	I	2
															25

Il dato principale che emerge dalla TAB. 7 immediatamente è che la quasi totalità delle SDEF riguarda il livello sintagmatico. Anche se, come abbiamo visto nel PAR. 5, la SDEF annotata al livello di sintagma coinvolge anche il livello di clausola (ad esempio accordo soggetto-verbo), non ci sono SDEF che possono definirsi esclusive solo del livello clausale (ad esempio inversioni di SP e SN), ed è per questo motivo che la TAB. 7 non riporta la colonna per il livello *clause*. C'è un caso di SDEF, invece, che può ascriversi solo al livello frasale ossia la costruzione *nonsense* vista nell'esempio (8).

Valutando i casi in cui le SDEF riguardano più livelli sintattici, si riporta nella TAB. 8 la distribuzione delle SDEF nei livelli sintattici che formano il testo.

TABELLA 8
Percentuali delle SDEF per livelli sintattici

	Frase	Clausola	Sintagma
SDEF %	10%	4%	86%

Sebbene la maggior parte delle SDEF riguardi soprattutto la costruzione sintagmatica, ovvero le relazioni grammaticali tra la testa del sintagma ed eventuali determinanti e modificatori, oppure la scelta lessicale della testa, alcune di esse coinvolgono anche livelli superiori come la clausola e in alcuni casi anche la frase (si veda l'esempio [II]).

Di sicuro ciò che emerge è che le SDEF che occorrono in un determinato livello linguistico non ostacolano la costruzione del livello sintattico superiore. Se si osservano i dati sulla quantità di frasi, clausole e sintagmi, si può vedere che la produzione dei pazienti e quelle dei soggetti di controllo non presentano differenze nella quantità di strutture sintattiche. Il numero totale delle clausole quasi coincide nei due gruppi di parlanti. Inoltre, anche la distribuzione delle clausole principali e subordinate non presenta notevoli differenze nei due sottocorpora. Lo stesso discorso vale per il rapporto tra il numero di clausole per ogni frase e il numero di sintagmi per ogni clausola.

TABELLA 9

Numero e percentuali delle strutture sintattiche nelle produzioni dei pazienti e dei soggetti di controllo

	Pazienti	Soggetti di controllo
Frasi	23	19
Clausole	76	73
Clausole principali	45%	42%
Clausole subordinate	50%	50%
Clausole senza verbo	5%	8%
Rapporto clausola/frase	3,3	3,8
Rapporto sintagma/clausola	2,2	2,4

Come si evince dalla TAB. 9, da un punto di vista quantitativo non vi sono notevoli differenze tra i due campioni di parlato. Sia nei testi prodotti dai pazienti sia in quelli dei soggetti di controllo le clausole principali sono poco meno della metà rispetto alle clausole subordinate. Allo stesso modo, non ci sono differenze per quanto riguarda la frequenza di frasi senza verbo.

La mancanza di differenze significative tra le due produzioni è confermata anche dal test di correlazione chi-quadro ($\chi^2 = 0.74987$, $df = 5$, $p\text{-value} = 0.9801$) e dal calcolo dei residui⁹, i cui risultati sono riportati nella TAB. 10.

TABELLA 10

Risultato del calcolo dei residui tra le frequenze dei tipi di unità sintattiche nelle produzioni dei parlanti afasici e del gruppo di controllo

	Pazienti	Soggetti di controllo
Frasi	0.29444014	-0.38936499
Clausole	0.13092367	-0.17313228
Clausole principali	0.15958933	-0.21103949
Clausole subordinate	-0.07828043	0.10351734
Clausole senza verbo	0.25296854	-0.33452333

Infine, un'ultima osservazione riguarda il rapporto tra il numero di parole delle descrizioni e il numero dei tipi di disfluenza: pause, pause piene o esitazioni (HES), cambiamenti di progettazione sintattica (RR) e ripetizioni (REP). I dati riportati nella TAB. 11 si riferiscono alla frequenza con cui ciascun fenomeno di disfluenza considerato occorre nei testi prodotti dai due gruppi di parlanti.

TABELLA II

Frequenza dei fenomeni di disfluenze per numero di parole registrate nei corpora dei pazienti e dei soggetti di controllo

	Pazienti	Soggetti di controllo
Numero di parole / HES	5,1	5,7
Num. Parole / RR	26,4	106,5
Num. Parole / REP	79,4	0

Per quanto riguarda la frequenza delle esitazioni, non ci sono differenze tra le produzioni dei pazienti e quelle dei soggetti di controllo, mediamente si producono pause circa ogni cinque parole. Per quanto riguarda le riformulazioni e i cambiamenti di progettazione sintattica, si registrano fenomeni di questo tipo circa ogni 26 parole nei testi dei pazienti e ogni 106 parole circa nelle descrizioni dei soggetti di controllo. Infine per quanto riguarda le ripetizioni, nei testi analizzati esse occorrono solo nelle produzioni dei parlanti afasici mediamente ogni 80 parole.

8 Conclusioni

Il sistema di annotazione manuale AN.ANA.S L2 ha permesso di esaminare la distribuzione delle SDEF del parlato afasico fluente di tre pazienti attraverso le categorie grammaticali e i livelli sintattici che costituiscono il testo. L'uso di etichette generali "omissione", "sostituzione" ecc. permette una maggiore estendibilità dell'analisi a diversi campi come quelli dell'apprendimento e del parlato afasico. Inoltre il suo schema di annotazione, che permette di rappresentare i rapporti gerarchici tra i livelli del testo annotato, ha consentito la valutazione delle strutture sintattiche prodotte dai pazienti, di annotare le strutture meno canoniche e i fenomeni di disfluenza frequenti nel parlato. Confrontando queste produzioni con i fenomeni della lingua orale il margine entro cui si fanno rientrare le strutture devianti o deficitarie si restringe e diventa specifico di questi testi (Kasselimis *et al.*, 2017; Nephyytou *et al.*, 2017).

Inoltre, nonostante le difficoltà che possono esserci nella descrizione univoca di una SDEF, attraverso AN.ANA.S è stato possibile annotare lo stesso fenomeno come appartenente contemporaneamente a più livelli sintattici.

Infine, l'utilizzo di un sistema di etichettatura valido per l'annotazione di testi prodotti da apprendenti e da parlanti afasici può favorire il confronto di dati di diversa provenienza. Come è noto, lo studio del linguaggio patologico può gettare luce sui meccanismi che riguardano l'acquisizione del linguaggio più in generale. Esistono infatti analogie tra le difficoltà linguistiche di individui affetti di afasie e quelle dei bambini che imparano a parlare su tutti i livelli linguistici del sistema grammaticale (Jakobson, 1971; 2002). Ad esempio, per quanto riguarda il lessico, le forme di anomia

differenziate per categorie semantiche o lessicali che caratterizzano le produzioni afasiche presentano tratti in comune con l’acquisizione delle parole nei primi anni di vita dei bambini (Basile, 2012a; 2012b).

Più in generale, lo studio dei processi che caratterizzano l’acquisizione e la perdita di diverse capacità linguistiche e comunicative può costituire un osservatorio di grande interesse per indagare i processi di denominazione e di costruzione di sensi e negoziazione di significati. I casi di acquisizione e di apprendimento linguistico da una parte, e di deficit linguistico dall’altra, possono rivelare aspetti fondamentali, regolarità «sui meccanismi coinvolti nell’elaborazione e nella produzione di strategie comunicative e scelte grammaticali» (Favilla, Nuzzo, 2015, p. 7).

Note

1. Il corpus è stato raccolto durante un periodo di tirocinio (febbraio-marzo 2011) presso l’ospedale San Camillo IRCCS di Venezia in collaborazione con l’Università Ca’ Foscari di Venezia sotto il coordinamento di Francesca Meneghelli e Anna Cardinaletti. Per il presente articolo ringrazio inoltre Riccardo Orrico per i preziosi consigli sull’interpretazione dei dati.

2. www.parlaritaliano.it.

3. Cfr. <https://aphasia.talkbank.org/protocol/>.

4. Cfr. <http://corpora.dslo.unibo.it/APHASIA/>.

5. Con *parlato afasico fluente* si definisce il parlato di persone con afasia di Wernicke o anche afasia sensoriale, afasia di conduzione, afasia transcorticale sensoriale e anomia (Edwards, 2005).

6. Corpus Ferrante (2016).

7. Cfr. <https://www.3c.org/XML>.

8. La sequenza potrebbe essere considerata un’inversione della forma *il suo*. Tuttavia si è ritenuto sufficiente considerarla come caso di aggiunta di materiale proprio perché i determinanti in questione introducono un nome di parentela (*fratello*) dove risulterebbe improbabile anche la forma *il suo fratello*, sebbene in quest’ultimo caso l’ordine dei determinanti fosse una forma target.

9. Valore di riferimento della significatività: -2.86526.

Bibliografia

- Andorno C., Grassi R., Valentini A. (2017), *Verso una nuova lingua: capire l’acquisizione di L2*, UTET Università, Novara.
- Armstrong E. (2000), *Aphasic discourse analysis: The story so far*, in “Aphasiology”, 14, 9, pp. 875-92.
- Basile G. (2012a), *Perdere le parole nei casi di afasia*, in A. Manco (a cura di), *Lessico e linguistica. Il lessico visto da vicino*, PUC, Cluj-Napoca, pp. 33-51.
- Basile G. (2012b), *La conquista delle parole. Per una storia naturale della denominazione*, Carracci, Roma.
- Basile G. (2014), *Perdere le parole. Disturbi linguistici tra afasia e schizofrenia*, Edizioni Accademiche italiane, Saarbrücken.
- Basile G. (2015), *Le parole per dir(si). Disfluenze e interventi di repair nei dialoghi del corpus CIPPS*, in F. Orletti, F. M. Dovetto, A. Cardinaletti, *Tra linguistica medica e linguistica clinica. Il ruolo del linguista*, in “Studi italiani di linguistica teorica e applicata”, 3, pp. 570-86.
- Basile G. (2017), *Le relazioni di significato nei dialoghi del corpus CIPPS*, in P. Sorianello (a cura di), *Il linguaggio disturbato. Modelli, strumenti, dati empirici*, Aracne, Roma, pp. 19-32.

- Bazzanella C. (1992), *Aspetti pragmatici della ripetizione dialogica*, in G. Gobber (a cura di), *La linguistica pragmatica*, Atti del xxiv Congresso della SLI, Bulzoni, Roma, pp. 433-54.
- Calzà L., Gagliardi G., Rossini Favretti R., Tamburini F. (2021), *Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia*, in "Computer speech and language", 65, pp. 1-18.
- Cutugno F., Voghera M. (2004), *AN.AN.A.S.: Analisi sintattica e annotazione XML a contatto*, in F. Alzano Leoni, F. Cutugno, M. Pettorino, R. Savo (a cura di), *Il parlato italiano. Atti del Convegno Nazionale*, D'Auria Editore, Napoli.
- De Mauro T. (2008), *In principio erat verbum*, in T. De Mauro, *Il linguaggio tra natura e storia*, Mondadori Università, Milano.
- Denes G. (2009), *Parlare con la testa: le basi neurologiche e la struttura del linguaggio*, Zanichelli, Bologna.
- Dovetto F. M. (a cura di) (2017), *Lingua e patologia. Le frontiere interdisciplinari del linguaggio*, Aracne, Roma.
- Dovetto F. M., Gemelli M. (2013), *Il parlar matto. Schizofrenia tra fenomenologia e linguistica. Il corpus CIPPS*, Aracne, Roma.
- Dovetto F. M., Panunzi P., Gregori L. (2017), *Sull'annotazione di un corpus orale mistilingue non standard (patologico schizofrenico)*, in A. De Meo, F. M. Dovetto (a cura di), *La comunicazione parlata*, Aracne, Roma, pp. 345-61.
- Edwards S. (2005), *Fluent aphasia*, Cambridge University Press, Cambridge.
- Favilla M. E., Ferroni L. (2018), *Il recupero linguistico nell'afasia post-ictus: apprendimento di una lingua seconda? Possibili interazioni tra afasiologia e linguistica acquisizionale*, in A. De Meo, M. Ralsuolo (a cura di), *Usare le lingue seconde. Comunicazione, tecnologia, disabilità, insegnamento*, Aitala, Milano, pp. 167-84, <http://www.aitla.it/images/pdf/StudiAItLA7/013Favilla-Ferroni.pdf>.
- Ferrante G. (2015-16), *Parlato e scritto a confronto. Una nuova metodologia*, Tesi di laurea magistrale, Università degli Studi di Salerno.
- Goodglass H., Kaplan E. (1983), *The assessment of aphasia and related disorders*, Lea e Febiger, Philadelphia, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1995234/>.
- Halliday M. A. K. (1985), *An introduction to functional grammar*, E. Arnold, London.
- Jakobson R. (1971), *Il farsi e il disfarsi del linguaggio: linguaggio infantile e afasia*, Giulio Einaudi Editore, Bologna, pp. 105-28.
- Jakobson R. (2002), *Due aspetti del linguaggio e due tipi di afasia*, in R. Jakobson, *Saggi di linguistica generale*, trad. it. di L. Heilmann e L. Grassi, Feltrinelli, Milano, pp. 22-45.
- Kassellimis D. S., Simos P. G., Peppas C., Evdokimidis I., Potagas C. (2017), *The unbridged gap between clinical diagnosis and contemporary research on aphasia: A short discussion on the validity and clinical utility of taxonomic categories*, in "Brain & Language", 164, pp. 63-7.
- Kassellimis D. S., Varkanitsa M., Angelopoulou G., Evdokimidis I., Goutsos D., Potagas C. (2020), *Word analysis in aphasia: Introducing the Greek Aphasia Corpus (GRAEC)*, in "Frontiers in Psychology", 11, 1577, <https://www.frontiersin.org>.
- Khudyakova M., Bergelson M., Akinina Y., Iskra E., Toldova S., Dragoy O. (2016), *Russian CLIPS: A corpus of narratives by brain-damaged individuals*, in D. Kokkinakis (ed.), *Proceedings of LREC 2016 workshop*, Linköping University Electronic Press, Portorož, pp. 22-6.
- Kong A. P., Law S. (2019), *Cantonese AphasiaBank: An annotated database of spoken discourse and co-verbal gestures by healthy and language-impaired native Cantonese speakers*, in "Behavior Research Methods", 51, pp. 1131-44.

- Kuvač Kraveljević J., Hržica G., Lice K. (2017), *CRODA: A croatian discourse corpus of speakers with aphasia*, in "Hrvatska revija za rehabilitacijska istraživanja", 2017, 2, 53, pp. 61-71.
- Laures-Gore J., Russell S., Patel R., Frankel M. (2016), *The Atlanta motor speech disorders corpus: Motivation, development, and utility*, in "Folia Phoniatrica Logopaedica", 68, pp. 99-105.
- Linnik A., Bastiaanse R., Höhle B. (2016), *Discourse production in aphasia: A current review of theoretical and methodological challenges*, in "Aphasiology", 30, 7, pp. 765-800.
- Luzzati C., Wilmes K., Bleser D. (1996), *AAT Aachner Aphasia Test* (edizione italiana), Giunti os, Firenze.
- MacWhinney B., Fromm D., Holland A., Forbes M. (2012), *AphasiaBank. Data and methods*, in N. Mueller, M. Ball (eds.), *Methods in clinical linguistics*, Wiley, New York, pp. 31-48.
- Miceli G., Laudanna A., Burani C., Capasso R. (1994), *Batteria per l'analisi dei deficit afasici. BADA*, CEPSAG, Roma.
- Mirman D., Strauss T. J., Brecher A., Walker G. M., Sobel P., Dell G. S. (2010), *A large, searchable, web-based database of aphasic performance on picture naming and other tests of cognitive function*, in "Cognitive neuropsychology", 27, 6, pp. 495-504.
- Neophytou K., van Egmond M., Avrutin S. (2017), *Zipf's law in aphasia across languages: A comparison of English, Hungarian and Greek*, in "Journal of Quantitative Linguistics", 24, 2-3, pp. 1-19.
- Nuzzo E., Favilla M. E. (2015), *Grammatica applicata: apprendimento, patologie, insegnamento*, Aitla, Officinaventuno, Milano, pp. 5-11, <http://www.aitla.it/images/pdf/eBook-AITLA-2.pdf>.
- Savy R. (2007), *Specifiche per la trascrizione ortografica annotata*, http://www.clips.unina.it/it/documenti/II_specifiche_trascrizione_ortografica.
- Selinker L. (1972), *Interlanguage*, in "International Revue of Applied Linguistics", 10, pp. 209-31.
- Thompson C. K., Shapiro L. P., Kiran S., Sobecks J. (2003), *The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of treatment efficacy (CATE)*, in "Journal of Speech, Language and Hearing Research", 46, pp. 591-607.
- Turco G., Voghera M. (2010), *From text to lexicon: The annotation of pre-target structures in an Italian learners corpus*, in M. Moneglia, A. Panunzi (eds.), *Bootstrapping Information from corpora in a cross-linguistic perspective*, Firenze University Press, Firenze, pp. 141-74.
- Varlokosta S., Stamouli S., Karasimos A., Markopoulos G., Kakavoulia M., Nerantzini M., Pantoula A., Fyndanis V., Economou A., Protopapas A. (2016), *A Greek corpus of aphasic discourse: Collection, Transcription, and annotation specifications*, in D. Kokkinakis (ed.), *Proceedings of LREC 2016 workshop*, Linköping University Electronic Press, Portorož.
- Voghera M. (2017), *Dal parlato alla grammatica*, Carocci, Roma.
- Voghera M., Basile G., Cerbasi D., Fiorentino G. (2004), *La sintassi della clausola nel dialogo*, in F. Albano Leoni, F. Cutugno, M. Pettorino, R. Savy (a cura di), *Il Parlato italiano. Atti del Convegno Nazionale (13-15.11.2003)*, M. D'Auria Editore, Napoli.
- Voghera M., Basile G., Cutugno F., Fiorentino G. (2005), *Sintassi in AN.ANA.S.*, in F. Albano Leoni, R. Giordano (a cura di), *Italiano Parlato. Analisi di un dialogo*, Liguori, Napoli, pp. 189-211.
- Westerhout E., Monachesi P. (2006), *A pilot study for a corpus of Dutch aphasic speech (CODAS)*, in *Proceedings of the Fifth International Conference on Language resources and evaluation (LREC 2006)*, pp. 1648-53.
- Williams C., Thwaites A., Buttery P., Geertzen J., Randall B., Shafto M. (2010), *The Cambridge Cookie-Theft Corpus: A corpus of directed and spontaneous speech of brain-damaged patients and healthy individuals*, in *Proceedings of the International Conference on Language resources and evaluation*, European Language Resources Association (ELRA), Valletta.