

- Sato, Sayaka / Gygax, Pascal M. / Gabriel, Ute. 2013. “Gender Inferences: Grammatical features and their impact on the representation of Gender in Bilinguals”, *Bilingualism: Language and Cognition* 16 (4): 792-807. <https://doi.org/10.1017/S1366728912000739>.
- Sato, Sayaka / Gabriel, Ute / Gygax, Pascal M. 2016. “Altering Male-Dominant Representations: A Study on Nominalized Adjectives and Participles in First and Second Language German”, *Journal of Language and Social Psychology* 35 (6): 667-685. <https://doi.org/10.1177/0261927X15625442>.
- Storme, Benjamin / Delaloye Saillen, Laura. 2022a. “Follow-up: Grammatical gender in French: Its impact on the interpretation of social gender in generic hybrid nouns” [Data set]. OSF. <https://doi.org/10.17605/OSF.IO/5XDBU>.
- Storme, Benjamin / Delaloye Saillen, Laura. 2022b. “Grammatical gender in French: Its impact on the interpretation of social gender in generic hybrid nouns” [Data set]. OSF. <https://doi.org/10.17605/OSF.IO/PRZVY>
- Thornton, Anna M. 2022. “Genere e igiene verbale: l’uso di forme con ə in italiano”, *AION-L*, no. 11: 11-54.
- Xiao, Hualin / Strickland, Brent / Peperkamp, Sharon. 2022. “How Fair Is Gender-Fair Language? Insights from Gender Ratio Estimations in French”, *Journal of Language and Social Psychology* 42 (1): 82-106. <https://doi.org/10.1177/0261927X221084643>.

WS3 – Linguistica teorica e trattamento automatico delle lingue: verso nuove sinergie

Arianna Bienati, Jennifer-Carmen Frey, Alessio Palmero Aprosio, Noemi Facchinelli

Applicazione delle risorse disponibili per l’italiano all’annotazione automatica delle relazioni discorsive in testi scolastici: alcune implicazioni teoriche.

L’annotazione delle relazioni discorsive gioca un ruolo fondamentale nell’analisi della struttura testuale, nonché nell’allenamento dei *discourse parser* che consentono di automatizzare queste analisi. Molti sono gli schemi di annotazione del discorso creati per l’inglese (cfr. *Penn Discourse Treebank*, Prasad et al. 2008; *Rhetorical Structure Theory Discourse Treebank*, Carlson / Marcu 2001; *Segmented Discourse Representation Theory*, Asher / Lascarides 2003), i quali sono stati successivamente utilizzati e adattati anche per altre lingue. Per l’italiano si ricordano gli studi di Pareti / Prodanoff (2010) e di Feltracco et al. (2017), i quali si concentrano su specifiche categorie di relazioni discorsive, rispettivamente quelle di attribuzione e quelle contrastive.

Una risorsa che comprende tutte le categorie del PDTB è il *Lexicon of Italian Connectives* (LICO; Feltracco et al. 2016), un lessico di connettivi italiani e i relativi sensi, allineato con la PDTB 3.0 (Webber et al. 2019) e dialogante con altri lessici simili disponibili per molte altre lingue (cfr. *Connective-Lex*; Stede et al. 2019). Il LICO si basa su una definizione di connettivo molto stretta, tratta dalle riflessioni teoriche di Ferrari (2010; 2014; 2021), che limita la categoria a quelle «forme linguistiche morfologicamente invariabili che offrono istruzioni su come legare gli eventi evocati dal testo o gli atti linguistici di composizione testuale attraverso relazioni logico-argomentative» (Ferrari 2021: 145). Tuttavia, i connettivi così intesi spesso non esauriscono tutti gli elementi formali che possono realizzare in superficie una relazione discorsiva. Per una risorsa come il LICO, il cui obiettivo è coadiuvare l'annotazione di relazioni discorsive, siano esse implicite o esplicite, è stato dunque importante trovare un compromesso: da una parte si cerca di essere fedeli alla definizione, escludendo forme linguistiche che esprimono la medesima relazione semantica ma appartengono a classi morfologicamente variabili (es. 'in conclusione', ma non 'concludendo' o 'per concludere'); dall'altra, si accettano espressioni multi-parola morfologicamente variabili (es. 'per [tutti] questo/i motivo/i') che però indubbiamente aumentano il potere descrittivo della risorsa.

Con il nostro contributo ci proponiamo di discutere alcune implicazioni teoriche della definizione di connettivo adottata in questa risorsa, portando a supporto delle nostre osservazioni i risultati di uno studio sulla coesione in testi di studenti delle scuole secondarie di primo e secondo grado. Utilizzando il LICO come base per l'identificazione delle relazioni discorsive, risponderemo alle seguenti domande di ricerca:

- a) in che modo le scelte in merito alla segmentazione in unità discorsive e la definizione di connettivo influenzano la ricostruzione e l'analisi della struttura logico-argomentativa di testi argomentativi scritti da studenti della secondaria di primo e secondo grado?
- b) in che misura queste scelte influenzano le prestazioni di un approccio automatico di identificazione dei connettivi basato sul semplice *match* di stringhe contenute nel LICO?

Per rispondere a queste domande, 30 testi (~10.000 token), di cui 18 tratti dalla componente italiana del corpus LEONIDE (Glaznieks et al. 2020) e 12 dal pre-test del progetto *ITACA – Coerenza nell'Italiano Accademico*, sono stati annotati manualmente da due annotatrici sulla base dello schema di annotazione della PDTB 3.0 (Webber et al. 2019). Il processo di annotazione si compone di due task: in primo luogo, si decide se l'elemento formale presente nel testo sia o meno un connettivo; in secondo luogo, si decide quale relazione discorsiva assegnare. Alla fine di questa procedura, i casi di mancato accordo sono stati discussi in una sessione di riconciliazione ed è stato concordato un *gold*

standard. L'annotazione manuale è stata poi comparata all'annotazione automatica, la quale consiste nel *match* delle stringhe contenute nel LICO con quelle contenute nei testi.

Nel riconoscimento dei connettivi, l'approccio automatico presenta una bassa *precision* (0.72): alcune stringhe (es. 'e', 'come', 'o', 'prima', 'dopo'), pur essendo invariabili, non connettono eventi con relazioni logico-argomentative nella maggior parte dei contesti di occorrenza (si pensi, ad esempio, alla 'e' che connette elementi nello stesso sintagma nominale), quindi non sono state segnalate come connettivi nell'annotazione manuale. Inoltre, a queste stesse stringhe risulta problematico assegnare una relazione discorsiva univoca, perché sono altamente polisemiche (ad esempio, il dizionario Sabatini-Coletti elenca almeno otto sensi per la 'e' congiuntiva). La *recall* dell'identificazione dei connettivi è 0.91, ma l'analisi delle differenze con l'annotazione manuale evidenzia il fatto che alcuni elementi che avrebbero sempre funzione connettiva non sono stati trovati, sia per errori ortografici sia perché esclusi a priori dal lessico, in quanto appartenenti a classi verbali (es. 'concludendo', 'visto che', 'dato che') oppure occorrenti in varietà non standard (es. 'essendo che').

I risultati suggeriscono che un approccio onomasiologico, cioè che parta dall'individuazione delle relazioni semantiche e solo successivamente passi al riconoscimento degli elementi formali che le segnalano, potrebbe favorire lo sviluppo di risorse sempre più precise per l'analisi automatica della struttura discorsiva per l'italiano. In particolare, si potrebbe attingere alle recenti sperimentazioni di Das / Taboada (2017; 2018; Das 2014) o Danlos (2018), in cui sono stati inventariati non solo i connettivi, ma anche altre strategie linguistiche (referenziali, lessicali, sintattiche o grafiche) che, in isolamento o in combinazione (Hoek et al. 2019), fungono da segnali per determinate relazioni discorsive. Un tale arricchimento delle risorse disponibili per l'italiano consentirebbe inoltre di sistematizzare le entrate già presenti nel lessico, distinguendo ciò che è connettivo, secondo la definizione rigorosa di Ferrari (2021), da ciò che non lo è, senza però diminuire il potere descrittivo della risorsa.

Riferimenti bibliografici

- Asher, Nicholas / Lascarides, Alex. 2003. *Logics of Conversation*. Cambridge, Cambridge University Press.
- Carlson, Lynn / Marcu, Daniel. 2001. *Discourse Tagging Reference Manual*. University of Southern California Information Sciences Institute. <https://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>

- Danlos, Laurence. 2018. "Discourse and Lexicons: Lexemes, MWEs, Grammatical Constructions and Compositional Word Combinations to Signal Discourse Relations". In Savary, Agata et al. (a cura di), *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018), Workshop at Coling 2018*. Santa Fe, Association for Computational Linguistics: 30-41. <https://hal.science/hal-02069442>.
- Das, Debopam. 2014. *Signalling of Coherence Relations in Discourse*. PhD thesis. Simon Fraser University.
- Das, Debopam / Taboada, Maite. 2014. *RST Signalling Corpus Annotation Manual*. Department of Linguistics, Simon Fraser University.
- Das, Debopam / Taboada, Maite. 2018. "RST Signalling Corpus: A corpus of signals of coherence relations", *Language Resources and Evaluation* 52(1): 149-184.
- Feltracco, Anna / Jezek, Elisabetta / Magnini, Bernardo / Stede, Manfred. 2016. "LICO: A Lexicon of Italian Connectives". In Corazza, Anna / Montemagni, Simonetta / Semeraro, Giovanni (a cura di), *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016*. Torino, Accademia University Press: 141-145.
- Feltracco, Anna / Magnini, Bernardo / Jezek, Elisabetta. 2017. "Contrast-Ita Bank: A corpus for Italian Annotated with Discourse Contrast Relations". In Basili, Roberto / Nissim, Malvina / Satta, Giorgio (a cura di), *Proceedings of the Fourth Italian Conference on Computational Linguistics CLiC-it 2017*. Torino, Accademia University Press: 159-164
- Ferrari, Angela. 2010. "Connettivi". In *Enciclopedia dell'Italiano Treccani*. Consultabile in rete all'indirizzo: [https://www.treccani.it/enciclopedia/connettivi_\(Enciclopedia-dell'Italiano\)](https://www.treccani.it/enciclopedia/connettivi_(Enciclopedia-dell'Italiano)).
- Ferrari, Angela. 2014. *Linguistica del testo. Principi, fenomeni, strutture*. Roma, Carocci.
- Ferrari, Angela. 2021. "Segnali discorsivi e connettivi", *Lingua e Stile* LVI(1): 143-150.
- Glaznieks, Aivars / Frey, Jennifer C. / Stopfner, Maria / Zanasi, Lorenzo / Nicolas, Lionel. 2022. "Leonide: A longitudinal trilingual corpus of young learners of Italian, German and English", *International Journal of Learner Corpus Research* 8(1): 97-120.
- Hoek, Jet / Zufferey, Sandrine / Evers-Vermeul, Jacqueline / Sanders, Ted J. M. 2019. "The linguistic marking of coherence relations: Interactions between connectives and segment-internal elements", *Pragmatics & Cognition* 25(2): 276-309.
- Pareti, Silvia / Prodanof, Irina. 2010. "Annotating Attribution Relations: Towards an Italian Discourse Treebank". In Calzolari, Nicoletta et al. (a cura di), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, European Language Resources Association (ELRA): 3566-3571.

- Prasad, Rashmi / Dinesh, Nikhil / Lee, Alan / Miltsakaki, Eleni / Robaldo, Livio / Joshi, Aravind / Webber, Bonnie. 2008. "The Penn Discourse Treebank 2.0". In Calzolari, Nicoletta et al. (a cura di), *Proceedings of the 6th International Conference of Language Resources and Evaluation (LREC 2008)*. Marrakech, European Language Resources Association (ELRA): 2961-2968.
- Stede, Manfred / Scheffler, Tatjana / Mendes, Amàlia. 2019. "Connective-Lex: A Web-Based Multilingual Lexical Resource for Connectives", *Discours* 24: 3-38
- Webber, Bonnie / Prasad, Rashmi / Lee, Alan / Joshi, Aravind. 2019. *The Penn Discourse Treebank 3.0 Annotation Manual*. Philadelphia, University of Pennsylvania.

Gaia Caligiore

La Lingua dei Segni Italiana (LIS) e la costruzione di un dataset in ambito medico: riflessioni teoriche

La costruzione di accessibilità attraverso l'interazione tra le tecnologie di elaborazione del linguaggio naturale e la Lingua dei Segni Italiana (LIS) richiede una riflessione epistemologica che tenga conto non solo delle peculiarità linguistiche della LIS, ma anche della prospettiva della comunità segnante e l'uso quotidiano che viene fatto della lingua. Infatti, l'applicazione di tecniche di elaborazione informatica alla LIS implica diverse sfide che devono essere affrontate per garantire l'efficacia e l'accuratezza delle soluzioni proposte. Dal punto di vista linguistico, queste sfide sono legate alla natura visivo-gestuale della lingua, che è multilineare e multimodale (Cuxac / Antinoro Pizzuto 2010), alla mancanza di una forma standardizzata di scrittura (Volterra et al. 2019), nonché alla limitata disponibilità di risorse (*corpora e dataset*).

Nello specifico, la traduzione in e dalla LIS richiede la creazione di un avatar o di uno scheletro che sia in grado di rappresentare accuratamente tutti i parametri formazionali che compongono il segnato, suddivisi in parametri manuali quali configurazione della mano, orientamento del palmo, luogo di esecuzione del segno e movimento, e in componenti corporee come l'espressione facciale, le componenti orali, il movimento del busto e la direzione dello sguardo (Volterra et al. 2019). Pertanto, al fine di generare una traduzione in LIS, è necessario, come per le lingue vocali, partire da processi di segmentazione e trascrizione dei segni, che verranno successivamente categorizzati per definirne relazioni e strutture. In questo contesto, tre questioni principali richiedono una riflessione: la descrizione degli enunciati LIS secondo strutture sintattiche fisse, la difficoltà riscontrata nel categorizzare i segni della LIS seguendo rigidi ruoli grammaticali e l'assenza di un sistema standardizzato di codifica per le lingue dei segni (LS).