

# Nome Cognome

Alessandro Mazzei

## Curriculum e pubblicazioni

### Principali aree d'interesse

Le principali aree di interesse di Alessandro Mazzei sono nel campo dell'intelligenza artificiale e più precisamente nell'area della linguistica computazionale. Esse si possono riassumere in tre macro aree: *Sintassi Computazionale*, *Semantica Computazionale*, *Ontologie Informatiche*.

### Principali Ricerche

All'interno degli approcci computazionali alla sintassi, Alessandro Mazzei si occupa principalmente di grammatiche formali, treebank e parsing.

Continuando la ricerca intrapresa nella sua tesi di dottorato, egli studia un nuovo formalismo grammaticale che sia formalmente appropriato alle caratteristiche delle lingue naturali e contemporaneamente anche psicolinguisticamente motivato. Questo progetto segue un approccio recente della linguistica computazionale, secondo cui i sistemi formali devono rispettare i modelli della linguistica e psicolinguistica in modo trasparente. L'idea fondamentale è di usare una grammatica formale che vincoli le produzioni solo attraverso la definizione delle strutture elementari e delle operazioni di base. Questo approccio si pone come un'alternativa al modus operandi di usare delle regole limitative su una grammatica formale computazionalmente meno vincolata. Alessandro Mazzei definisce un formalismo grammaticale che estende le grammatiche "Tree Adjoining" e che modella la natura incrementale del linguaggio umano. Questo progetto è motivato da una crescente mole di dati sperimentali provenienti dalla psicolinguistica, e dai suggerimenti di alcune teorie linguistiche, che congetturano che gli esseri umani analizzino le parole delle frasi in maniera incrementale. Questa modellizzazione dei meccanismi coinvolti nell'elaborazione del linguaggio comporta una separazione più sfumata tra la "competence" e la "performance" delle grammatiche, in quanto la definizione delle strategie di elaborazione (es. da sinistra verso destra) è parte integrante del formalismo [2, 7, 19].

Un'altra linea di ricerca sviluppata da Alessandro Mazzei nell'ambito della sintassi computazionale riguarda l'estrazione automatica di grammatiche da dati annotati sintatticamente, i cosiddetti treebank. La creazione di grandi treebank, come il "Penn Treebank" per l'inglese, ha portato all'ideazione di nuove tecniche automatiche per l'estrazione di grammatiche di ampia copertura. Basandosi sulla notazione sintattica presente nel "Turin University Treebank", un treebank a dipendenze per l'italiano scritto, Alessandro Mazzei ha studiato ed implementato una procedura che permette di estrarre una grammatica "tree adjoining" per l'italiano. Questo strumento ha consentito di verificare alcune misure sulla complessità del parsing per le grammatiche "Tree Adjoining", e su come la copertura di tali grammatiche sia influenzata dalla natura dei dati da cui vengono estratte [6]. Sempre nel campo della sintassi computazionale rientra l'attività svolta da Alessandro Mazzei nelle competizioni EVALITA'07 e EVALITA'091 : EVALITA è una serie di eventi dedicati alla valutazione degli strumenti per il trattamento automatico della lingua italiana. In questi progetti Alessandro Mazzei si è occupato dell'organizzazione e della valutazione del task riguardante il parsing [5, 10, 12, 16, 17].

Alessandro Mazzei ha lavorato recentemente anche nel campo delle semantica computazionale. La digitalizzazione di grandi quantità di documenti ha reso disponibile una grande quantità di informazione di vario tipo in formato testuale. In generale, i sistemi automatici basati sulla linguistica computazionale possono essere usati per estrarre dell'informazione semantica da questi documenti. Alessandro Mazzei si è occupato di costruire degli analizzatori semantici per estrarre automaticamente informazioni da documenti legislativi. In particolare egli si è interessato di modifiche normative, cioè di leggi che modificano altre leggi: partendo dalla struttura XML di questi documenti, egli ha studiato il problema dell'inserimento automatico di meta-dati, che caratterizzano l'effetto delle modifiche legali contenute nel documento [11, 20, 18]. Questa attività di ricerca è stata sviluppata nel contesto dei progetti di ricerca I fondamenti cognitivi per la creazione del diritto e ICT4LAW.

Un altro lavoro di ricerca a cavallo tra sintassi e semantica computazionale a cui Alessandro Mazzei ha partecipato riguarda la trasformazione del "Turin University Treebank" in un insieme di derivazioni categoriali. Le grammatiche categoriali hanno un'interfaccia diretta tra sintassi e semantica e questo in prospettiva permetterà di convertire il Turin University Treebank in un treebank semantico [15].

Alessandro Mazzei ha lavorato anche nel campo delle ontologie informatiche. In particolare egli si è occupato della progettazione e della creazione di un'ontologia legale. Esiste oramai una necessità reale di condivisione dell'informazione legale nell'ambito dell'Unione Europea: con questa motivazione, Alessandro Mazzei ha partecipato allo sviluppo di un sistema ontologico che permettesse l'annotazione delle direttive europee2 rispetto alle varie legislazioni nazionali. Queste ricerche hanno portato alla realizzazione di un prototipo software, chiamato "Legal Taxonomy Syllabus", che mediante un'architettura web di tipo client-server, permette l'annotazione ontologica delle direttive da parte dei diversi gruppi di studio delle nazioni partecipanti al progetto [1, 4, 3, 8, 9, 14, 13]. Questa attività di ricerca è stata sviluppata nel contesto dei progetti di ricerca Uniform Terminology Project e ICT4LAW.

### **Principali Pubblicazioni**

- [1] L. Lesmo, G. Boella, A. Mazzei, and P. Rossi. Multilingual Conceptual Dictionaries Based on Ontologies: Analytical Tools and Case Studies. In *Proc. of Conference "Approaching the Multilanguage Complexity of European Law: Methodologies in Comparison"*, pages 1–14, Florence, November 2006. ITTIG.
- [2] A. Mazzei and V. Lombardo. Towards a dynamic constituency model of syntax. In *Proc. of XL International Congress of Società Linguistica Italiana, SLI06*, pages 1–15, Vercelli, October 2006.
- [3] G. Ajani, G. Boella, L. Lesmo, A. Mazzei, and P. Rossi. Multilingual ontological analysis of european directives. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 21–24, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [4] G. Ajani, L. Lesmo, G. Boella, A. Mazzei, and P. Rossi. Terminological and ontological analysis of european directives: multilinguism in law. In *11-th International Conference on Artificial Intelligence and Law, ICAIL07*, pages 43–48, Stanford, California, June 2007.
- [5] C. Bosco, A. Mazzei, and V. Lombardo. Evalita parsing task: an analysis of the first parsing system contest for italian. *Intelligenza Artificiale*, 2(4):30–33, September 2007.
- [6] A. Mazzei and V. Lombardo. Building a wide coverage Dynamic Grammar. *Applied Artificial Intelligence*, 21(4-5):281–296, April 2007.

- [7] A. Mazzei, V. Lombardo, and P. Sturt. Dynamic TAG and Lexical Dependencies. *Research on Language and Computation*, 5(3):309–332, March 2007.
- [8] G. Ajani, G. Boella, L. Lesmo, A. Mazzei, D.P. Radicioni, and P. Rossi. Legal Taxonomy Syllabus: Handling Multilevel Legal Ontologies. In *Proceedings of Langtech 2008*, Roma, February 2008.
- [9] G. Ajani, G. Boella, L. Lesmo, A. Mazzei, D.P. Radicioni, and P. Rossi. Multilevel legal ontologies. In *International Conference on Language Resources and Evaluation*, Marrakech, Marocco, May 2008.
- [10] C. Bosco, A. Mazzei, V. Lombardo, G. Attardi, A. Corazza, A. Lavelli, L. Lesmo and G. Satta, and M. Simi. Comparing italian parsers on a common treebank: the evalita experience. In *Proc. of 6th International Conference on Language Resources and Evaluation, LREC08*, Marrakech, Marocco, May 2008.
- [11] R. Brighi, L. Lesmo, A. Mazzei, M. Palmirani, and D.P. Radicioni. Towards semantic interpretation of legal modifications through deep syntactic analysis. In *Jurix 2008: The 21st Annual Conference, Frontiers in Artificial Intelligence and Applications*. IOS Press, December 2008.
- [12] B. Magnini, A. Cappelli, F. Tamburini, C. Bosco, A. Mazzei, V. Lombardo, F. Bertagna, N. Calzolari, A. Toral, V. Bartalesi Lenzi, R. Sprugnoli, and Manuela Speranza. Evaluation of natural language tools for italian: Evalita 2007 evaluation of natural language tools for italian: Evalita 2007. In *Proc. of 6th International Conference on Language Resources and Evaluation, LREC08*, Marrakech, Marocco, May 2008.
- [13] G. Ajani, G. Boella, L. Lesmo, M. Martin, A. Mazzei, M. Martin, D.P. Radicioni, and P. Rossi. Multilevel Legal Ontologies. In S. Montemagni, editor, *Semantic Processing of Legal Texts*, Berlin, December 2009. Springer.
- [14] G. Ajani, G. Boella, L. Lesmo, M. Martin, A. Mazzei, D.P. Radicioni, and P. Rossi. Legal Taxonomy Syllabus version 2.0. In *3rd Workshop on Legal Ontologies and Artificial Intelligence Techniques*, Barcelona, Spain, June 2009. CEUR.
- [15] J. Bos, C. Bosco, and A. Mazzei. Converting a Dependency-Based Treebank to a Categorical Grammar Treebank for Italian. In *Proceedings of the eighth International workshop on Treebanks and Linguistic Theories - TLT8*, Milano, December 2009.
- [16] C. Bosco, A. Mazzei, and V. Lombardo. Evalita'09 Parsing Task: constituency parsers and the Penn format for Italian. In *Proceedings of Evalita'09*, Reggio Emilia, December 2009.
- [17] C. Bosco, S. Montemagni, A. Mazzei, V. Lombardo, F. Dell'Orletta, and A. Lenci. Evalita'09 Parsing Task: comparing dependency parsers and treebanks. In *Proceedings of Evalita'09*, Reggio Emilia, December 2009.
- [18] L. Lesmo, A. Mazzei, and D.P. Radicioni. Semantic Annotation of Legal Modificatory Provisions. In R. Serra, editor, *AI\*IA 2009: 10th Congress of the Italian Association for Artificial Intelligence*, LNAI, Berlin, December 2009. Springer-Verlag.
- [19] A. Mazzei, V. Lombardo, and P. Sturt. Constraining the form of supertags with the strong connectivity hypothesis. In S. Bangalore and A. Joshi, editors, *Complexity of Lexical Descriptions and its Relevance to Natural Language Processing: A Supertagging Approach*. MIT Press, December 2009. To appear.
- [20] A. Mazzei, D.P. Radicioni, and R. Brighi. NLP-based Extraction of Modificatory Provisions Semantics. In *Proceedings of the International Conference on Artificial Intelligence and Law, ICAIL09*, Barcelona, Spain, June 2009. ACM.