

**‘Specialization in ICTs and Special Education: Psychopedagogy of Integration’  
Postgraduate Program  
DEMOCRITUS UNIVERSITY OF THRACE Department of Greek Philology  
in collaboration with  
NCSR DEMOKRITOS Informatics and Telecommunications Institute**

**« CREATING A TEXT CORPUS OF SPEAKING CHILDREN WITH  
AUTISTIC SPECTRUM DISORDERS (ASD) »**

MAZIOTI GEORGIA

POSTGRADUATE  
THESIS

SUPERVISORY COMMITTEE

1. Γαβριηλίδου Ζωή  
Καθηγήτρια Τ.Ε.Φ. Δ.Π.Θ.
2. Μητσιάκη Μαρία  
Επίκουρη Καθηγήτρια Τ.Ε.Φ. Δ.Π.Θ.
3. Δρίγκας Αθανάσιος  
Διευθυντής Ερευνών & Ερευνητής Α' βαθμίδας Ι.Π.Τ. Ε.Κ.Ε.Φ.Ε. «ΔΗΜΟΚΡΙΤΟΣ»

KOMOTINI / ATHENS 2018

# « Creating a Text Corpus of Speaking Children with Autistic Spectrum Disorders (ASD) »

**Keywords:** text corpus, Autistic Spectrum Disorders (ASD), oral speech, Greek language, child Language

## Abstract

This diploma thesis is placed in the field of computational linguistics, in which are included text corpora. In particular, the thesis focuses on the basic concepts concerning the definition and types of text corpora. It also attempts to present the existing corpus in Greek language focusing mainly on child language corpora. In addition, the present work describes the selection criteria and organization of corpora content and their annotation process.

The aim of this thesis is to design a special text corpus that includes transcripts from the Greek speaking children diagnosed with Autism Spectrum Disorders (ASD). The creation of this corpus aims to fill the void that exists in Greek concerning child language corpora and create a text corpus that will be the beginning for the study of oral speech of children with ASD.

## References

- Abbeduto, L., Seltzer, M.M., Shattuck, P., Krauss, W.M., Orsmond, G. & Murphy, M.M. (2004). Psychological Well-Being and Coping in Mothers of Youths With Autism, Down Syndrome, or Fragile X Syndrome. *American Journal on Mental Retardation*, 109 (3), 237 - 254.
- Adolphs, S. (2008). *Corpus and Context. Investigating pragmatic functions in spoken discourse*. Amsterdam: John Benjamins.
- Adolphs, S. & Carter, R. (2013). *Spoken Corpus Linguistics: From Monomodal to Multimodal*. London: Routledge.
- Adolphs, S. & Knight, D. (2010). Building a spoken corpus. What are the basics? In A. O'Keefe & M. McCarthy (Ed.), *The Routledge Handbook of Corpus Linguistics* (pp. 38 - 52). London: Routledge.
- Aijmer, K. & Altenberg, B. (Ed.). (1991). *English Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4<sup>th</sup> ed., text rev.)*. Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, DC: American Psychiatric Association.
- Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistics Research*, 30 (2), 141 - 161.
- Asperger, M. (1991). Autistic psychopathy in childhood. In U. Frith (Ed). *Autism and Asperger's Syndrome*. Cambridge: Cambridge University Press.
- Aston, G., Bernardini, S. & Stewart, D. (2004). *Corpora and Language Learners*. Amsterdam/Philadelphia: Benjamins.
- Atkins, S., Clear, J., Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7 (1), 1 - 16.

- Bauer, M. W. & Aarts, B. (2000). Corpus construction: A principle for qualitative data collection. In M. W. Bauer, & G. Gaskell (Ed.), *Qualitative Researching with Text, Image and Sound* (pp. 19 - 37). London: Sage.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8 (4), 243 - 257.
- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Βογινδρούκας, Ι. (1999, Μάρτιος). Αυτισμός - Διαταραχή επικοινωνίας. *Επικοινωνιολόγος, φωνή ομιλία*, 10.
- Bol, G.W. & Kuiken, F. (1990). Grammatical analysis of developmental language disorders: A study of the morphosyntax of children with specific language disorders, with hearing impairment and with Down's syndrome. *Clinical Linguistics and Phonetics*, 4, 77 - 86.
- Bowker, L. & Pearson, J. (2002). *Working with Specialized Language: A practical Guide to Using Corpora*. London: Routledge.
- Γαβριηλίδου, Ζ. & Καμπάκη-Βουγιούκλη, Π. (2011). Φωνολογική περιγραφή του Greek-speaking Children Corpus (GSCC). Στο Κ. Χατζοπούλου, & Α. Ιωαννίδου (Επιμ.), *Πρακτικά 9ου Διεθνές Συνεδρίου Ελληνικής Γλωσσολογίας (ICGL)*, (σσ. 559 - 568).
- Γαβριηλίδου, Μ., Λαμπροπούλου, Π. & Ρονιώτη, Σ. (1993). Σχεδιασμός και σχολιασμός ενός ελληνικού σώματος κειμένων. *Μελέτες για την ελληνική γλώσσα*, 308 – 322, Θεσσαλονίκη.
- Γαβριηλίδου, Μ., Χριστοφίδου, Α. & Λαμπροπούλου, Π. (1994). Σύστημα Κατηγοριοποίησης Γραπτών Κειμένων. *Μελέτες για την Ελληνική Γλώσσα: Πρακτικά της 15ης συνάντησης του Τομέα Γλωσσολογίας, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης*. Θεσσαλονίκη.

- Γενά, Α. (2002). *Αυτισμός και διάχυτες αναπτυξιακές. Αξιολόγηση - διάγνωση - αντιμετώπιση*. Αθήνα: Leader Books.
- Γιάγκου, Μ. (2010). Ηλεκτρονικά Σώματα Κειμένων για τη διδασκαλία της ελληνικής γλώσσας: Επισκόπηση της βιβλιογραφίας και των υπαρχόντων Σωμάτων Κειμένων. *i-teacher (Περιοδικό για τις Τ.Π.Ε. και τις εφαρμογές τους στην εκπαίδευση)*, σσ. 55 - 65.
- Conti-Ramsden, G. & Dykins, J. (1991). Mother-child interactions with language-impaired children and their siblings. *British Journal of Disorders of Communication*, 26, 337 - 354.
- Γούτσος, Δ. (2002). Η χρήση των ηλεκτρονικών σωμάτων κειμένων στην ανάλυση λόγου. Στο C. Clairis (επιμ.). *Γλωσσολογικές έρευνες για την ελληνική*. Ι. Πρακτικά του 5ου Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας, Σορβόνη, Παρίσι, 13-15 Σεπτεμβρίου 2001. Paris: L'Harmattan, 219-222.
- Γούτσος, Δ. (2003). Σώμα Ελληνικών Κειμένων: Σχεδιασμός και υλοποίηση. Πρακτικά του 6ου Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας, Πανεπιστήμιο Κρήτης, 18-21 Σεπτεμβρίου 2003. Διαθέσιμο στο: <http://www.philology.uoc.gr/conferences/6thICGL/gr.htm> [Τελευταία πρόσβαση: 5 Μαρτίου 2018].
- Γούτσος, Δ. (2014). Συνδέοντας το κείμενο με το σύστημα: Γλωσσικές πράξεις και σώματα κειμένων. Στο Σ. Μ. Γεωργαλίδου, *Ανάλυση λόγου: Θεωρία και εφαρμογές* (σσ. 189 - 223). Αθήνα: Νήσος.
- Γούτσος, Δ. & Φραγκάκη, Γ. (2014). Πρόσφατη γλωσσική αλλαγή στα ελληνικά : Σχεδιασμός του διαχρονικού σώματος ελληνικών κειμένων του 20ου αιώνα. *Πρακτικά του 11ου Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας* (σσ. 318 - 329). Ρόδος: Πανεπιστήμιο Αιγαίου.
- Γούτσος, Δ. & Φραγκάκη, Γ. (2015). *Εισαγωγή στη γλωσσολογία σωμάτων κειμένων*. Αθήνα: Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράματα και Βοηθήματα, Διαθέσιμο στο: [www.kallipos.gr](http://www.kallipos.gr) [Πρόσβαση 12 Ιανουαρίου 2018].

- Dipper, S. (2008). Theory-driven and corpus-driven computational linguistics and the use of corpora. In A. Lüdeling, & M. Kyto, *Corpus Linguistics: An International Handbook* (pp. 68 - 96). Berlin: Walter de Gruyter.
- Folstein, S. & Rutter, M. (1977). Infantile autism: a genetic study of 21 twin pairs. *Journal of clinical of psychology and psychiatry*, 18, 297 - 321.
- Frith, U. (1999). *Αυτισμός, εξηγώντας το αίνιγμα*. (Γ. Καλομοίρης, Μεταφρ.) Αθήνα: Ελληνικά Γράμματα.
- Georgakopoulou, A. & Goutsos, D. (1998). Conjunctions versus discourse markers in Greek: The interaction of frequency, positions and functions in context. *Linguistics*, 36 (5), 887 - 917.
- Gillberg, I.C. & Gillberg, C. (1989, July). Asperger syndrome - some epidemiological considerations: a research note. *Journal of Child Psychology and Psychiatry*, 30 (4), 631-638.
- Goutsos, D., Hatzidaki, R. & King, Ph. (1994). A corpus-based approach to Modern Greek language research and teaching. In I. Philippaki-Warburton, K. Nicolaidis, & M. Sifianou (Ed.), *Themes in Greek Linguistics: Papers from the First International Conference on Greek Linguistics* (pp. 507 - 513). Amsterdam/Philadelphia: John Benjamins.
- Goutsos, D., King, Ph. & Hatzidaki, R. (1994). Towards a Corpus of Spoken Modern Greek. *Literary and Linguistic Computing*, 9 (3), 215 - 223.
- Halladay, K. A., Bishop, S., Constantino, J. N., Daniels, A. M., Koenig, K., Palmer, K., Messinger, D., Pelphrey, K., Sanders, S. J., Singer, A. T., Taylor, A. L. & Szatmari, P. (2014). Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. *Molecular Autism*, 6 (1), 36.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

- Jordan, R. (2000). *Η εκπαίδευση παιδιών και νεαρών ατόμων με αυτισμό*. (Ι. Καφαντάρης, Μεταφρ.) Αθήνα: Ελληνικά Γράμματα.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 217 - 250.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London - New York: Longman.
- Κυπριωτάκης, Α. (2003). *Τα αυτιστικά παιδιά και η αγωγή τους*. Αθήνα: Γρηγόρη.
- Leech, G. (1997). Introducing corpus annotation. In R. Garside, G. Leech, & A. McEnery, *Corpus Annotation* (Ed.), (pp. 1 - 18). London: Longman.
- Lüeling, A. & Kyto, M. (2009). *Corpus Linguistics - An International Handbook* (Vol. 2). Berlin - New York: Walter de Gruyter.
- Luzon, M. J. & Campoy, M. C. (2007). *Spoken Corpora in Applied Linguistics (Linguistic Insights, Studies in Language and Communication)*. Peter Lang AG.
- Μαγνήσαλης, Σ. & Σαγιά, Σ. (2014). *DSM – official site / NAS / Autism Speaks / Ilona Roth (Understanding Autism in the 21st century)*. Νότας, Σ. Εκδ..
- McCarthy, M. & O'Keeffe, A. (2010). Historical perspective. What are corpora and how have they evolved? In A. O'Keeffe, & M. McCarthy (Ed.), *The Routledge Handbook of Corpus Linguistics* (pp. 3 - 13). London: Routledge.
- McEnery, T. & Wilson, A. (1996). *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Μαρκόπουλος, Γ. Α. (2006). *Ζητήματα υπολογιστικής γλωσσολογίας: Prolog και μορφολογική ανάλυση*. Αθήνα: Περιοδικό "Παρουσία" - Παράρτημα 69.
- Mesibov, G., Adams, L. & Klinger, L. (1997). *Autism, Understanding the disorder*. Plenum Press (48).
- Μικρός, Γ. (2004). Ηλεκτρονικά σώματα κειμένων και ορολογία. Στο Κ. Μ. Ελένη (Επιμ.), *Ελληνική ορολογία: έρευνες και εφαρμογές* (σσ. 139 - 164). Αθήνα: Καστανιώτη.

- Παπαντωνίου, Μ. & Καμπούρογλου, Μ. (2003). *Ανάπτυξη και Διαταραχές της επικοινωνίας και του λόγου στον Αυτισμό*. Διαθέσιμο στο: [http://www.epsyme.gr/admin/files/2\\_Epikoinonia\\_Epimorfosi.pdf](http://www.epsyme.gr/admin/files/2_Epikoinonia_Epimorfosi.pdf) [Τελευταία Πρόσβαση: 5 Μαρτίου 2018].
- Πιπερίδης, Σ. & Παπαγεωργίου, Χ. (1995). Υπολογιστική επεξεργασία σωμάτων κειμένων. Στο *Μελέτες για την ελληνική γλώσσα* (σσ. 775 - 786). Θεσσαλονίκη.
- Πολίτης, Π. (n.d.). *Βελτίωση και επέκταση των σωμάτων κειμένων της Ηλεκτρονικής Πύλης του Κέντρου Ελληνικής Γλώσσας (ΚΕΓ)*. Διαθέσιμο στη: δικτυακή πύλη για την Ελληνική Γλώσσα: [http://www.greek-language.gr/greekLang/modern\\_greek/bibliographies/corpora/pyli.html](http://www.greek-language.gr/greekLang/modern_greek/bibliographies/corpora/pyli.html) [Τελευταία πρόσβαση: 15 Απριλίου 2018].
- Πολίτης, Π. (n.d.). *Ο Εθνικός Θησαυρός Ελληνικής Γλώσσας (ΕΘΕΓ) του Ινστιτούτου Επεξεργασίας του Λόγου (ΙΕΛ)*. Διαθέσιμο στη: δικτυακή Πύλη για την Ελληνική Γλώσσα: [http://www.greek-language.gr/greekLang/modern\\_greek/bibliographies/corpora/ethel.html](http://www.greek-language.gr/greekLang/modern_greek/bibliographies/corpora/ethel.html) [Τελευταία Πρόσβαση: 15 Απριλίου 2018].
- Πολίτης, Π. (n.d.). *Το σώμα κειμένων του Ηλεκτρονικού Κόμβου του Κέντρου Ελληνικής Γλώσσας (ΚΕΓ)*. Διαθέσιμο στη: δικτυακή Πύλη για την Ελληνική Γλώσσα: <http://www.greek-language.gr/greekLang/bibliographies/corpora/komvos.html> [Τελευταία Πρόσβαση: 15 Απριλίου 2018].
- Quill, Κ. (2000). *Do-Watch-Listen-Say: Social and Communication Intervention for Children with Autism*. Baltimore, United States: Brookes Publishing Co.
- Rollins, Ρ. (1999). Pragmatic accomplishments and vocabulary development in pre-school children with autism. *American Journal of Speech-Language Pathology: A Journal of Clinical Practice*, 8, 85 - 94.
- Rutter, Μ. (1990). *Νηπιακός αυτισμός. Σύγχρονες αντιλήψεις και αντιμετώπιση (Μτφρ.)*. Αθήνα: Ελληνικά Γράμματα.

- Rutter, M. (1990). Psychosocial resilience and protective mechanisms. In J. Rolf, A. Masten, D. Cicchetti, K. Nuechterlein, & S. Weintraub (Ed.), *Risk and protective factors in the development of psychopathology*. Cambridge: New York.
- Rutter, M. & Schopler, E. (1992). Classification of Pervasive developmental disorders: Some concepts and considerations. *Journal of Autism and Developmental Disorders*, 22, 459 - 482.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Statkiewicz-Gayhard, V., Peerenboom, B. & Campell, R.N. (2001). Διασχίζοντας τις γέφυρες. Η γονεϊκή προοπτική στην αντιμετώπιση ενός παιδιού αφού έχει διαγνωστεί με Αυτισμό/Διάχυτη αναπτυξιακή Διαταραχή. (Α. Λαζαρίνη, Μεταφρ.) Αθήνα: ΕΕΠΑΑ.
- Stephany, U. (1997). The acquisition of Greek. (D. I. Slobin, Ed.) *The Crosslinguistic Study of Language Acquisition*, 4, pp. 183-333.
- Stubbs, M. (2001). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Φρανσίς, Κ. Γ. (2012). *Η προβολή των διαταραχών του αυτιστικού φάσματος (ΔΑΦ) στην ενήλικη ζωή*. Αθήνα, Β' Ψυχιατρική Κλινική, ΕΚΠΑ: Πανεπιστημιακό Γενικό Νοσοκομείο "Αττικόν".
- Φραντζή, Κ. Θ. (2012). *Εισαγωγή στην επεξεργασία σωμάτων κειμένων*. Αθήνα: Ίων.
- Wing, L. (1981). Asperger's Syndrome: A clinical account. *Psychological Medicine*, 11, 115 - 129.
- Wing, L. (2000). *Το αυτιστικό φάσμα: Ένας οδηγός για τους γονείς και επαγγελματίες για τη διάγνωση*. Αθήνα: ΕΕΠΑΑ.
- World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

Wynne, M. (2005). *Developing Linguistics Corpora: A Guide to Good Practice*. Oxford: Oxbow Books.

Χαλτσιανή, Ι. (12 - 14 Δεκεμβρίου 2008). Ηλεκτρονικά σώματα κειμένων και γλωσσική διδασκαλία: Διεθνείς αναζητήσεις και διαφαινόμενες προοπτικές για την ελληνική γλώσσα. *Πρακτικά Συνεδρίου "2008, Ευρωπαϊκό Έτος Διαπολιτισμικού Διαλόγου: συνομιλώντας με τις γλώσσες - πολιτισμούς"* (σφ. 818 - 830). Θεσσαλονίκη: Τμήμα Γαλλικής Γλώσσας και Φιλολογίας ΑΠΘ.

Χαλτσιανή, Ι. (2009). Εκδοχές αξιοποίησης των ηλεκτρονικών κειμένων της νέας ελληνικής στη διδασκαλία της μητρικής γλώσσας. *Πρακτικά Πανελληνίου Συνεδρίου με διεθνή συμμετοχή "Η Διδασκαλία της Ελληνικής Γλώσσας (ως πρώτης/μητρικής, δεύτερης/ξένης)"*. Φλώρινα: Πανεπιστήμιο Δυτικής Μακεδονίας.

Χατζηγεωργίου, Ν., Σπηλιωτοπούλου, Α., Βακαλοπούλου, Α., Παπακωστοπούλου, Α., Πιπερίδης, Σ., Γαβριηλίδου, Μ. & Καραγιάννης, Γ. (2001). Εθνικός Θησαυρός Ελληνικής Γλώσσας (ΕΘΕΓ): Σώμα Κειμένων της Νέας Ελληνικής στο Διαδίκτυο. *Μελέτες για την Ελληνική Γλώσσα: Πρακτικά της 21ης συνάντησης του Τομέα Γλωσσολογίας του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης*, (σφ. 812 - 821).