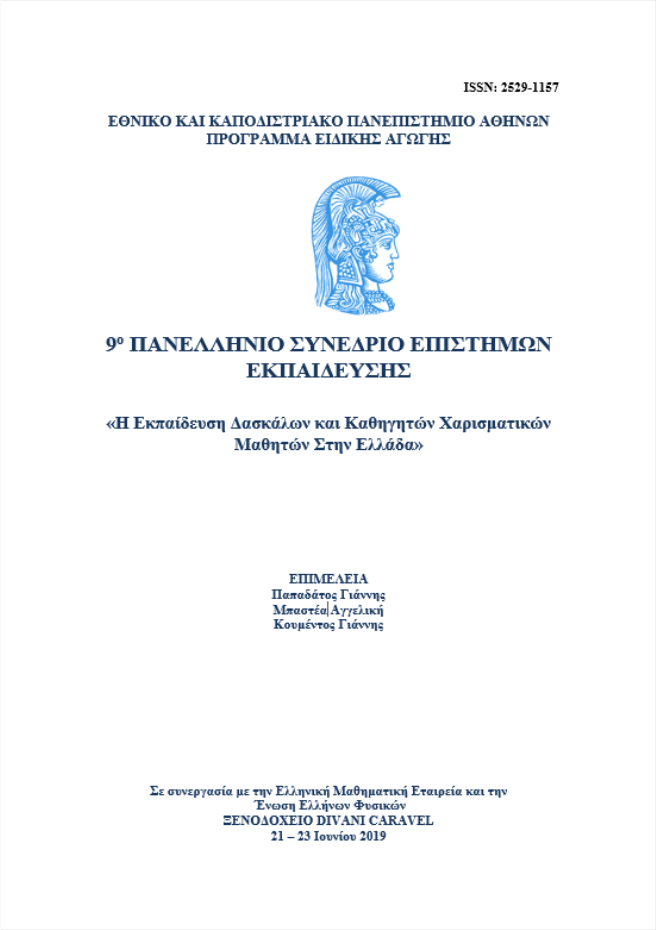
ePublishing
ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ
NATIONAL DOCUMENTATION CENTRE

Πανελλήνιο Συνέδριο Επιστημών Εκπαίδευσης

Τόμ. 9 (2019)

9ο ΠΑΝΕΛΛΗΝΙΟ ΣΥΝΕΔΡΙΟ ΕΠΙΣΤΗΜΩΝ ΕΚΠΑΙΔΕΥΣΗΣ



ISSN: 2529-1157

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΠΡΟΓΡΑΜΜΑ ΕΙΔΙΚΗΣ ΑΓΩΓΗΣ

9ο ΠΑΝΕΛΛΗΝΙΟ ΣΥΝΕΔΡΙΟ ΕΠΙΣΤΗΜΩΝ ΕΚΠΑΙΔΕΥΣΗΣ

«Η Εκπαίδευση Δασκάλων και Καθηγητών Χαρισματικών Μαθητών Στην Ελλάδα»

ΕΠΙΜΕΛΕΙΑ
Παπαδάτος Γιάννης
Μπαστέα|Αγγελική
Κουμέντος Γιάννης

Σε συνεργασία με την Ελληνική Μαθηματική Εταιρεία και την Ένωση Ελλήνων Φυσικών
ΞΕΝΟΔΟΧΕΙΟ DIVANI CARAVEL
21 – 23 Ιουνίου 2019

«Δημιουργία corpus γραπτού και προφορικού λόγου παιδιών με δυσορθογραφία και αυτισμό»

Γεωργία Μαζιώτη, Μαργαρίτα Τσεβά

doi: 10.12681/edusc.3176

Βιβλιογραφική αναφορά:

Μαζιώτη Γ., & Τσεβά Μ. (2020). «Δημιουργία corpus γραπτού και προφορικού λόγου παιδιών με δυσορθογραφία και αυτισμό». *Πανελλήνιο Συνέδριο Επιστημών Εκπαίδευσης*, *9*, 837–850. https://doi.org/10.12681/edusc.3176

«Δημιουργία corpus γραπτού και προφορικού λόγου παιδιών με δυσορθογραφία και αυτισμό»

Γεωργία Μαζιώτη
Ειδική Παιδαγωγός
gwgwmaz90@hotmail.com

Μαργαρίτα Τσεβά
Ειδική Παιδαγωγός
margaret-tse@hotmail.com

Περίληψη

Τα τελευταία χρόνια έχει αναπτυχθεί στη βιβλιογραφία η μελέτη και η έρευνα για τη δημιουργία σωμάτων κειμένων (corpora) με ιδιαίτερη έμφαση στην ανάπτυξη ειδικών corpora παιδαγωγικού σκοπού. Η παρούσα μελέτη έρχεται να συμβάλει σε αυτό το πλαίσιο και ιδιαίτερα στην ανάγκη μελέτης της χρήσης της γλώσσας των παιδιών με αυτισμό ή ειδικές μαθησιακές δυσκολίες, λαμβάνοντας υπόψη τη συμβολή των σωμάτων κειμένων στη γλωσσολογική έρευνα, τις προκλήσεις σχεδιασμού τους αλλά και τη δυναμικότητα του πεδίου των Ηλεκτρονικών Σωμάτων Κειμένων (ΗΣΚ). Πιο συγκεκριμένα, αξιοποιώντας τόσο θεωρητικά εργαλεία όσο και εργαλεία σχεδιασμού επιχειρεί να μελετήσει τους όρους σχεδιασμού ενός ειδικού γραπτού σώματος κειμένων με δείγμα κείμενα με ορθογραφικά λάθη παιδιών με ειδικές μαθησιακές δυσκολίες (ΕΜΔ - δυσορθογραφία) και ενός ειδικού προφορικού σώματος κειμένων με δείγμα απομαγνητοφωνημένα κείμενα στα ελληνικά από προφορικό λόγο παιδιών που έχουν διαγνωστεί με Διαταραχές Αυτιστικού Φάσματος (ΔΑΦ). Επιπλέον, στόχος της είναι η δημιουργία ενός ηλεκτρονικού σώματος κειμένων για εκπαιδευτικούς σκοπούς σε θεωρητικό επίπεδο που θα καλύψει το κενό που υφίσταται στα ελληνικά όσον αφορά τα σώματα κειμένων παιδικής γλώσσας. Μέσα από τη συγκεκριμένη μελέτη διαφαίνεται ότι ο σχεδιασμός σωμάτων κειμένων αποτελεί ένα χρήσιμο εργαλείο κατηγοριοποίησης και εύκολης πρόσβασης των λαθών που κάνουν τα παιδιά τόσο για τους εκπαιδευτικούς όσο και για κάθε ενδιαφερόμενο.

Λέξεις–Κλειδιά: σώμα κειμένων; Διαταραχές Αυτιστικού Φάσματος (ΔΑΦ); Δυσορθογραφία; προφορικός/γραπτός λόγος; παιδική γλώσσα

Abstract

Recent studies on the creation of corpora have emphasized the need for the development of special corpora for pedagogical purposes. The purpose of the present study is to contribute to this development by looking into the language of children with autism or language disorders. More specifically, the study aims at setting the theoretical framework for the creation of an electronic corpus for educational purposes. To pursue this objective, the study includes readings from text objects and transcriptions from spoken language into linguistic research, their design challenges as well as the potential of the field of electronic corpora by making use of the WordSmith Tools application and the resulting online corpus.

The main research question of the study is whether the corpus methodology is useful both for understanding miscellaneous spelling and phonological mistakes and for encoding them. It is argued that a corpus design is a useful tool for teachers, so they can assess the mistakes made by children. Through the out-coming results the present study attempts to become the stimulus for filling the research gap in Greek literature

regarding the development of children's textbooks with dysorthography and children's spoken language with autism and, as such, to be a valuable educational intervention tool.

Keywords: corpora; autism; dysorthography; spoken/written language; children's language

ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ

1. Ορισμός σωμάτων κειμένων

Σύμφωνα με τον Sinclair (1991, 1996), τα σώματα κειμένων (corpus, πληθ. corpora) ορίζονται ως μια «συλλογή τμημάτων γλώσσας, τα οποία επιλέγονται και διατάσσονται σύμφωνα με συγκεκριμένα γλωσσολογικά κριτήρια, έτσι ώστε να μπορούν να χρησιμοποιηθούν ως αντιπροσωπευτικό δείγμα μιας συγκεκριμένης γλώσσας» ή μια «συλλογή κομματιών γλωσσικού κειμένου σε ηλεκτρονική μορφή, που επιλέγονται σύμφωνα με εξωτερικά κριτήρια, ώστε να αντιπροσωπεύουν, στον βαθμό που είναι δυνατόν, μια γλώσσα ή μια γλωσσική ποικιλία ως πηγή δεδομένων για τη γλωσσολογική έρευνα» (Sinclair, 2005). Ο όρος γλώσσα δεν περιορίζεται εδώ στην έννοια μόνο της εθνικής γλώσσας αλλά δηλώνει και τη γλωσσική ποικιλία ή την ειδική γλώσσα.

Συγκεκριμένα, τα σώματα κειμένων παρέχουν στους γλωσσολόγους τη δυνατότητα για επιστημονικές παρατηρήσεις σχετικά με τη συχνότητα χρήσεως ενός τύπου, το συμφραστικό περιβάλλον του, καθώς και —στη διαχρονική γλωσσολογική ανάλυση— σχετικά με την πρώτη εμφάνιση ή σημασία ενός τύπου και τις μετέπειτα αλλαγές του. Η έκταση ενός corpus ποικίλλει ανάλογα με τις πηγές και τον σκοπό για τον οποίο έχει παραχθεί, ενώ ειδικά σώματα κειμένων αντλούν επιλεκτικά στοιχεία από έντυπο ή προφορικό λόγο, με σκοπό τη μελέτη ορισμένης γλωσσικής λειτουργίας. Επομένως, βασικό στοιχείο ενός σώματος κειμένων είναι το γεγονός ότι δεν είναι απλά μια συλλογή πληροφοριών, αλλά μια αναπαράσταση της γλώσσας και της ποικιλίας της γλώσσας (Kennedy, 1998).

Με το πέρασμα των χρόνων οι υπολογιστές έδωσαν νέα πνοή στην λεξικογραφία καθιστώντας δυνατή και πιο εύκολη τη συλλογή και αποθήκευση μεγάλου όγκου πραγματικού γλωσσικού υλικού. Πλέον, η επεξεργασία με ηλεκτρονικά σώματα κειμένων είναι γρηγορότερη και πιο αξιόπιστη διότι οι υπολογιστές έχουν τη δυνατότητα να επεξεργάζονται μεγάλο όγκο γλωσσικού υλικού με περισσότερη ακρίβεια (Μικρός, 2004).

2. Υπάρχοντα σώματα κειμένων

Μελετώντας τη βιβλιογραφία έχουν αναπτυχθεί ειδικά σώματα κειμένων παιδικής γλώσσας. Παραδείγματα αποτελούν: α) το **Corpus Stephany**, το οποίο διεξήχθη από την Ursula Stephany στο Πανεπιστήμιο της Κολωνίας στη Γερμανία μεταξύ 1971 και 1974 και περιλαμβάνει διαλόγους ανάμεσα σε τέσσερα μονόγλωσσα παιδιά που μεγάλωσαν στην Αθήνα και τις μητέρες ή τις γιαγιάδες τους (Stephany, 1997) και β) το **Greek Speaking Children Corpus** (GSCC) (Gavriilidou, & Kambaki - Vougioukli, 2011), το οποίο περιλαμβάνει καταγραφές προφορικού αυθόρμητου λόγου παιδιών από 3 – 6 χρονών και έχει ως στόχο τη μελέτη της φωνολογικής εκμάθησης και τον εντοπισμό φωνολογικών λαθών.

Παράλληλα, με την δημιουργία σωμάτων κειμένων παιδικής γλώσσας έχουν δημιουργηθεί ειδικά σώματα κειμένων μη τυπικής ανάπτυξης της παιδικής γλώσσας όπως: α) το **Spanish Corpus of Dyslexic Texts,** που πρόκειται για ένα

επισημειωμένο corpus μελέτης των λαθών που κάνουν τα παιδιά που έχουν μητρική γλώσσα τα ισπανικά και που έχουν διαγνωστεί με δυσλεξία στο γραπτό λόγο (Rello et al., 2012), β) το **Arabic Dyslexia Corpus (BDAC)**, το οποίο είναι ένα επισημειωμένο σώμα κειμένων και το περιεχόμενό του αποτελείται από τα λάθη ατόμων με δυσλεξία στο γραπτό λόγο. Η μητρική τους γλώσσα των συμμετεχόντων είναι τα αραβικά (Alamri & Teahan, 2017), γ) το **Croatian Error - Annotated Corpus of Non-Professional Written Language,** που σκοπός του είναι να συγκρίνει τα γλωσσικά λάθη υγιών ατόμων με εκείνων που έχουν διαγνωστεί με γλωσσικές διαταραχές στο γραπτό λόγο. Παράλληλα, δίνει έμφαση στην κοινωνική ένταξη ατόμων με γλωσσικές διαταραχές (Štefanec & al., 2016), δ) το **Conti-Ramsden 2 Corpus,** το οποίο δημιουργήθηκε από την Gina Conti-Ramsden στην Μεγάλη Βρετανία. Στο corpus αυτό συμμετείχαν τρία παιδιά με Ειδική Γλωσσική Διαταραχή και τα τρία τυπικής ανάπτυξης και αδέρφια μεταξύ τους (Conti-Ramsden, G. & Dykins, J., 1991), ε) το **Bol / Kuiken Corpus: SLI,** που πραγματοποιήθηκε από τους Gerard Bol και Folkert Kuiken στην Ολλανδία μεταξύ 1984 και 1988. Περιλαμβάνει δεδομένα από τα 20 παιδιά με ΕΓΔ με μη επαρκή πνευματική λειτουργία και μέσο δείκτη νοημοσύνης (Bol, G.W. &Kuiken, F., 1990) και στ) το **Rollins** Corpus, το οποίο δημιουργήθηκε από την Pamela Rosenthal Rollins και υποστηρίχθηκε από το Αμερικανικό Ίδρυμα Ομιλίας και Ακοής. Περιλαμβάνει μεταγραφές 5 αγοριών με αυτισμό που φοιτούν σε πρόγραμμα προσχολικής αγωγής ειδικό σχολείο στο Πανεπιστήμιο του Τέξας στο Ντάλας (Rollins, 1999).

3. Είδη και κατηγορίες σωμάτων κειμένων

Στη βιβλιογραφία (Sinclair, 1991; McEnery & Wilson, 1996; Kennedy, 1998; Hunston, 2002; Μικρός, 2004; McEnery & Xiao, 2007; Γούτσος & Φραγκάκη, 2015) έχουν διατυπωθεί διάφορα κριτήρια για τη διάκριση διαφορετικών ειδών σωμάτων κειμένων. Τα βασικότερα από αυτά περιλαμβάνουν (α) *την επιλογή της γλώσσας*, (β) *το εύρος της γλωσσικής κάλυψης*, (γ) *την τροπικότητα*, (δ) *τη χρονική κάλυψη*, (ε) *την ανανεωσιμότητα των δεδομένων* και (στ) *το είδος της επισημείωσης*. Τα σώματα κειμένων, λοιπόν, διαφοροποιούνται ανάλογα με τα παραπάνω κριτήρια και μπορεί να λάβουν έναν πιο γενικό ή ειδικό χαρακτήρα πέρα από τη μορφή ή το μέσο με το οποίο εμφανίζονται (Kennedy, 1998), ενώ η κατηγοριοποίησή τους δεν είναι αυστηρή αλλά δυναμική (Γούτσος & Φραγκάκη, 2015). Επισημαίνεται, πάντως, πως κατά τη διάρκεια του σχεδιασμού ενός σώματος κειμένων τα κριτήρια σύμφωνα με τα οποία σχεδιάζεται ένα corpus παραμένουν σταθερά κατά τη διάρκεια της έρευνας (McEnery & Xiao, 2007).

Ανάλογα, λοιπόν, με τον **αριθμό των γλωσσών** που περιέχουν τα σώματα κειμένων διακρίνονται σε ***μονόγλωσσα και πολύγλωσσα.*** Αντίστοιχα, τα πολύγλωσσα σώματα κειμένων, στον βαθμό που περιλαμβάνουν περισσότερες από μία γλώσσες ή/και διαλέκτους, υποκατηγοριοποιούνται σε ***συγκρίσιμα*** (comparable corpora) ή ***παράλληλα*** (parallel corpora) με κριτήριο την αναλογία των διαφορετικών κειμενικών ειδών (Μικρός, 2004; McEnery & Xiao, 2007, Γούτσος & Φραγκάκη, 2015).

Ανάλογα με το **εύρος των γλωσσικών ποικιλιών** που καλύπτουν τα σώματα κειμένων διακρίνονται σε:

- ***γενικά σώματα κειμένων*** (general corpora), τα οποία περιέχουν γραπτό και προφορικό λόγο, εξυπηρετούν γενικούς σκοπούς, ως κύριο στόχο θέτουν την αναπαράσταση της φυσικής γλώσσας ή υπόγλωσσας και παράλληλα προσφέρουν στοιχεία για μια γλώσσα στο σύνολό της και

- ***ειδικά σώματα κειμένων*** (specialized corpora), τα οποία χρησιμοποιούνται όταν θέλουμε να αναφερθούμε σε συγκεκριμένες χρήσεις της γλώσσας, καθώς εξειδικεύονται σε μια από τις παραμέτρους της γλωσσικής ποικιλότητας**.** Υποκατηγορία των ειδικών corpora είναι τα ***σώματα κειμένων μάθησης*** (learner corpora), τα οποία τα συναντάμε ως επί το πλείστον στον εκπαιδευτικό τομέα και περιλαμβάνουν προφορικά ή γραπτά κείμενα μη φυσικών ομιλητών μιας γλώσσας (Kennedy, 1998).

Με βάση το κριτήριο της **τροπικότητας** τα σώματα κειμένων διακρίνονται σε:

- ***προφορικού λόγου*** (spoken corpora), τα οποία είναι αρκετά απαραίτητα στους γλωσσολόγους, αφού βοηθούν στην έρευνα για τα χαρακτηριστικά του προφορικού λόγου γενικά, την προφορά και τις διαλέκτους μιας γλώσσας. Χαρακτηριστικό είναι ότι είναι πιο δύσκολο να συγκροτηθούν τα προφορικά σώματα κειμένων διότι προϋποθέτουν την ηχογράφηση ή καταγραφή προφορικών δεδομένων και συνήθως τη μεταγραφή τους σε γραπτό λόγο,
- ***γραπτού λόγου*** (written corpora), όπου περιέχουν αποκλειστικά και μόνο γραπτά κείμενα και
- ***μικτά***, καθώς περιέχουν κείμενα γραπτού και προφορικού λόγου (Γούτσος & Φραγκάκη, 2015).

Σύμφωνα με τη **χρονική κάλυψη** και τη χρονική περίοδο που καλύπτουν τα corpora κατηγοριοποιούνται σε:

- ***συγχρονικά*** (synchronic corpora), αφού περιλαμβάνουν στοιχεία μιας γλώσσας τα οποία έχουν κοινό σημείο αναφοράς, δηλαδή αναφέρονται σε μια συγκεκριμένη χρονική περίοδο και
- ***διαχρονικά*** (diachronic corpora), λόγω του ότι περιλαμβάνουν κείμενα της ίδιας γλώσσας ή γλωσσικής ποικιλίας που περιγράφουν μια μεγάλη χρονική περίοδο (Mair & al., 2002; Γούτσος & Φραγκάκη, 2014).

Όσον αφορά το κριτήριο της **ανανεωσιμότητας** τα corpora διακρίνονται σε:

- ***ανοικτά σώματα κειμένων ή σώματα κειμένων ελέγχου*** (monitor corpora), τα οποία είναι συνεχώς εξελισσόμενα σώματα, καθώς εμπλουτίζονται με νέα στοιχεία και ανανεώνονται συνεχώς κρατώντας σταθερή τη σύσταση των κειμενικών τους ειδών και
- ***κλειστά σώματα κειμένων***, που δεν επιδέχονται καμία περεταίρω αναθεώρηση αφότου ολοκληρωθεί η διαδικασία σύνταξής τους. Αυτά τα σώματα κειμένων είναι στατικά και έχουν δεδομένο μέγεθος και σύσταση έτσι ώστε να αποτελούν σταθερές και αμετάβλητες πηγές δεδομένων μετά την ολοκλήρωσή τους (Γούτσος & Φραγκάκη, 2015).

Τέλος, σχετικά με το κριτήριο της **επισημείωσης** (annotation), δηλαδή τη διαδικασία προσθήκης ερμηνευτικών γλωσσικών πληροφοριών, ένα corpus μπορεί να είναι:

α) ***μη επισημειωμένο*** (un-annotated ή untagged), όταν δεν έχει υποστεί αλλαγή σε γραμματικό, συντακτικό και σημασιολογικό επίπεδο τότε το σώμα κειμένου ή

β) ***επισημειωμένο*** όταν έχει προστεθεί κάποια γραμματική, συντακτική ή σημασιολογική πληροφορία (Γούτσος & Φραγκάκη, 2015).

4. Προσδιορισμός περιεχομένου – Κριτήρια οργάνωσης και κατηγοριοποίησης

Σύμφωνα με τον ορισμό ενός corpus σημαντικός στόχος στη συγκρότηση του είναι η επιδίωξη της αντιπροσωπευτικότητας (representativeness) και πιο συγκεκριμένα η επιλογή ενός αντιπροσωπευτικού δείγματος της υπό μελέτης

γλώσσας (Γουτσος, 2003). Σύμφωνα με τον Leech (1997), ένα σώμα κειμένων θεωρείται αντιπροσωπευτικό δείγμα μιας γλώσσας όταν τα ευρήματα που βασίζονται στο περιεχόμενό του μπορούν να γενικευτούν για το σύνολο της γλώσσας αυτής. Για το λόγο αυτό έχουν γίνει προσπάθειες καθορισμού κριτηρίων και αρχών για την κατάρτιση των σωμάτων κειμένων. Συνοπτικά, οι πιο βασικές αρχές επίτευξης της αντιπροσωπευτικότητας είναι η συνοχή, το μέγεθος, η ποικιλία ειδών κειμένων/ δειγματοληψία και οι ποικιλίες της γλώσσας (Sinclair, 2005).

Η κατηγοριοποίηση των κειμένων μπορεί να γίνει με αρκετούς τρόπους. Ο σημαντικότερος είναι αυτός που υιοθετεί το πρότυπο PAROLE (PAROLE 1995), το οποίο βασίζεται στις οδηγίες και αρχές του TEI (Text Encoding Initiative) και του προγράμματος EAGLES. Τα κείμενα, λοιπόν, ταξινομούνται σύμφωνα με το Μέσο Δημοσίευσης-επικοινωνιακό μέσο, το Κειμενικό Είδος και το Θέμα/Περιεχόμενο και την Ημερομηνία Παραγωγής τους (Barnard & Ide, 1995; Sperger & Burnard, 1994).

ΕΡΕΥΝΗΤΙΚΟ ΜΕΡΟΣ

5. Σώμα κειμένων γραπτού λόγου παιδιών με Ειδικές Μαθησιακές Δυσκολίες – Δυσορθογραφία

5.1 Ορισμός-Χαρακτηριστικά ΕΜΔ

Οι Μαθησιακές Δυσκολίες μπορούν να οριστούν ως μια «καθυστέρηση ή διαταραχή της ανάπτυξης σε μία ή περισσότερες λειτουργίες του γραπτού ή του προφορικού λόγου (ανάγνωση, γραφή, ορθογραφία, κατανόηση) ή και των μαθηματικών, εξαιτίας κάποιας πιθανής εγκεφαλικής δυσλειτουργίας ή διαταραχών συμπεριφοράς και συναισθημάτων» (Kirk,1962). Συγκεκριμένα, οι Μαθησιακές Δυσκολίες κατηγοριοποιούνται σε:

- *Γενικές Μαθησιακές Δυσκολίες (ΓΜΔ), που* αναφέρονται σε γενικότερες διαταραχές στο γνωστικό σύστημα του ατόμου (Πόρποδας, 2002) και
- *Ειδικές Μαθησιακές Δυσκολίες (ΕΜΔ), οι οποίες* σχετίζονται με τη μη επαρκή ανάπτυξη ειδικών λεκτικών, γλωσσικών, κινητικών και σχολικών δεξιοτήτων λόγω ενδογενών νευροφυσιολογικών παραγόντων και όχι άλλων αναπηριών ή διαταραχών (American Psychiatric Association, 2000). Η ελληνική νομοθεσία κατηγοριοποιεί τις Ειδικές Μαθησιακές Δυσκολίες με βάση τη γλωσσολογική τους έκφραση στη *δυσλεξία*, τη *δυσγραφία*, τη *δυσαριθμησία*, τη *δυσαναγνωσία* και τη *δυσορθογραφία*.

Αναλύοντας τη δυσορθογραφία που αποτελεί το βασικό πεδίο έρευνας του γραπτού corpus, μπορούμε να αναφέρουμε πως σύμφωνα με τη βιβλιογραφία πρόκειται για μια δυσκολία στο γραπτό λόγο και μπορεί να οριστεί γενικά ως δυσκολία απόκτησης της ικανότητας της ακολουθίας ενός συστήματος γραφής και πιο συγκεκριμένα να πάρει μορφές έλλειψης ικανότητας σύνδεσης μίας λέξης με σύμβολα ή εικόνες, έλλειψης ικανότητας σύνδεσης φωνημάτων και αντίστοιχων γραφικών συμβόλων ή έλλειψης κατανόησης της σχέσης μεταξύ ηχητικών και γραπτών ακολουθιών.

5.2 *Σχεδίαση Βάσης Δεδομένων*

Αρχικά, σύμφωνα με τα είδη και τις κατηγορίες των σωμάτων κειμένων που έχουν αναλυθεί στο θεωρητικό μέρος, το παρόν σώμα κειμένων έχει κάποια συγκεκριμένα χαρακτηριστικά. Συγκεκριμένα, πρόκειται για ένα **μονόγλωσσο**, **ειδικό** σώμα κειμένων **γραπτού λόγου**, που περιλαμβάνει την παραγωγή γραπτού λόγου παιδιών διαγνωσμένων με ειδικές μαθησιακές δυσκολίες και πιο

συγκεκριμένα με δυσορθογραφία. Παράλληλα, το συγκεκριμένο corpus είναι **συγχρονικό** λόγω του ότι τα δεδομένα που περιλαμβάνει έχουν συλλεχθεί κατά τη διάρκεια της σχολικής χρονιάς 2017 – 2018, **ανανεώσιμο** αφού υπάρχει η δυνατότητα εμπλουτισμού του με νέα στοιχεία κρατώντας σταθερή τη σύσταση του και επισημειωμένο, προκειμένου να γίνεται αντιληπτή η γραμματική κατηγορία ή/και η συντακτική θέση των λέξεων. Η εισαγωγή των κειμένων πραγματοποιείται με τη βοήθεια ενός σαρωτή ή μέσω της δακτυλογράφησης με το χέρι σε περιπτώσεις που υπάρχει δυσκολία ανάγνωσης και διάκρισης των γραμμάτων.

Σε αυτό το σημείο είναι χρήσιμο να προσδιορίσουμε τις αρχές δημιουργίας και σχεδιασμού του παρόντος σώματος κειμένων.

- **Περιεχόμενο**: ορθογραφικά λάθη παιδιών διαγνωσμένων με Ειδικές Μαθησιακές Δυσκολίες - δυσορθογραφία.
- **Σκοπός**: να αποτελέσει μια σαφώς καθορισμένη συλλογή γραπτού λόγου παιδιών με ΕΜΔ και ένα εργαλείο αυτόματου εντοπισμού και μελέτης των ορθογραφικών λαθών στα ελληνικά παιδαγωγικά αξιοποιήσιμο και μακροπρόθεσμα ο σχεδιασμός ενός κατάλληλου προγράμματος παρέμβασης.
- **Δημιουργός / Συγγραφέας**: 2 παιδιά ηλικίας 10-14 ετών διαγνωσμένα με ΕΜΔ – δυσορθογραφία.
- **Μέγεθος**: 8.000 λέξεις
- **Αριθμός κειμένων**: 146
- **Γνωστικό πεδίο**: Γλώσσα, Κοινωνική και Πολιτική Αγωγή, Θρησκευτικά, Φυσική, Μαθηματικά, Ιστορία
- **Δειγματοληψία**: γραπτά κείμενα που συλλέχθηκαν από τα τετράδια και τα βιβλία των συμμετεχόντων.
- **Κειμενικά Είδη**: α) κείμενα αυθόρμητης παραγωγής γραπτού λόγου, β) κείμενα που έχουν αντιγράψει και γ) κείμενα που έχουν παραχθεί καθ' υπαγόρευση.
- **Αντιπροσωπευτικότητα**: τα κείμενα είναι πολλά σε αριθμό, έχουν παραχθεί με φυσικό τρόπο, βρίσκονται σε άμεση συνάρτηση με τους στόχους μας και υπάρχει η δυνατότητα εμπλουτισμού του corpus.
- **Τήρηση μεταδεδομένων**: τα κείμενα έχουν κωδικοποιηθεί και περιέχουν πληροφορίες για τη σχέση των κειμένων με τους συμμετέχοντες στην έρευνα και για την ταξινόμηση των κειμένων ανάλογα με το είδος, το παιδί και τα ορθογραφικά λάθη.
- **Γλώσσα**: Ελληνική.
- **Ημερομηνία έκδοσης**: σχολικό έτος 2017 – 2018.
- **Διαθεσιμότητα**: ελεύθερα διαθέσιμο στην ιστοσελίδα του εργαστηρίου Γλωσσολογίας +ΜορΦωΣη του Τμήματος Ελληνικής Φιλολογίας του Δ.Π.Θ. http://synmorphose.compulaw.gr.

Τα ορθογραφικά λάθη κατηγοριοποιούνται σε: α) φωνολογικού τύπου λάθη όπως παραλείψεις, αντικαταστάσεις, προσθέσεις και αντιστροφές γραμάτων ή συλλαβών, μη σωστή χρήση κεφαλαίων καθώς και συνδυασμούς φωνημάτων που αντιστοιχούν σε ένα γράφημα, β) λάθη στο θέμα της λέξης, γ) λάθη στις καταλήξεις λέξεων που η ορθογραφία τους βασίζεται σε γραμματικούς κανόνες, δ) έλλειψη ή λαθεμένη χρήση των σημείων στίξης, ε) τονισμό (παρατονισμός) ή παντελή έλλειψη τόνων, στ) απουσία απόστασης μεταξύ λέξεων στην πρόταση, ζ) κακή γραφή των γραμμάτων, η οποία παρατηρείται κυρίως όταν υπαγορεύεται το κείμενο, ενώ σε περιπτώσεις γραφής των κειμένων στο σπίτι η γραφή των γραμμάτων είναι καλή.

Τέλος, κάθε κείμενο που αποτελεί το δείγμα του παρόντος corpus έχει κωδικοποιηθεί με έναν συγκεκριμένο αύξοντα αριθμό με σκοπό τη διάκριση του, ενώ έχει επισημειωθεί σύμφωνα με το μοντέλο PAROLE σε α) μορφολογικό επίπεδο, όπου τα λήμματα φέρουν πληροφορίες λημματολογίου (π.χ. σύνδεση με άλλα λήμματα, ορθογραφικές εκδοχές, κτλ.) και πληροφορίες μορφολογίας (γραμματική κατηγορία και υποκατηγορία, κλιτικό παράδειγμα, θέματα) και β) συντακτικό επίπεδο, όπου η λέξη κωδικοποιείται με τη μορφή συντακτικών μονάδων.

6. Σώμα κειμένων προφορικού λόγου παιδιών με Διαταραχές Αυτιστικού Φάσματος

6.1 Ορισμός-Χαρακτηριστικά ΔΑΦ

Οι Διαταραχές του Αυτιστικού Φάσματος (ΔΑΦ) είναι μια ομάδα νευροαναπτυξιακών διαταραχών που διαγιγνώσκονται στην πρώιμη παιδική ηλικία, αλλά επιμένουν σε όλη τη ζωή, αν και σημαντικές αλλαγές μπορούν να συμβούν. Απαντώνται σε συχνότητα 1–1,2% και κυμαίνονται από ηπιότερες, με ελάχιστα και ηπιότερα αυτιστικά στοιχεία, μέχρι βαρύτερες μορφές, με πολλαπλά αυτιστικά στοιχεία συνοδευόμενα από βαριά νοητική καθυστέρηση. Η συγκεκριμένη αναπηρία εμποδίζει τα άτομα να κατανοούν σωστά όσα βλέπουν, ακούν και γενικά αισθάνονται. Αυτό έχει ως αποτέλεσμα να αντιμετωπίζουν σοβαρά προβλήματα στις κοινωνικές σχέσεις, στην επικοινωνία και στη συμπεριφορά τους. Είναι το αποτέλεσμα μιας νευρολογικής διαταραχής που επηρεάζει τη λειτουργία του εγκεφάλου και εκδηλώνεται με άτυπη μορφή επικοινωνίας, κοινωνικών δεξιοτήτων και ικανοτήτων για παιχνίδι και ανταποκρίσεων στις εισερχόμενες αισθητηριακές πληροφορίες (Statkiewicz Gayhardt et al., 2001). Ο όρος «αυτισμός» προέρχεται από το ελληνικό «αυτός» που σημαίνει «εγώ ο ίδιος». Αυτή τη στιγμή για το Φάσμα του Αυτισμού χρησιμοποιούνται 2 διαγνωστικά συστήματα: το DSM-V (Diagnostic and Statistical Manual) της Αμερικανικής Ψυχιατρικής Ένωσης (American Psychiatric Association, 2013) και το ICD (International Classification of Diseases) του Διεθνούς Οργανισμού Υγείας (World Health Organization, 1992).

Στα πλαίσια της αυτιστικής διαταραχής η γλώσσα δεν είναι παραγωγική και δημιουργική και παρουσιάζει δυσκολίες στην κατανόηση. Γενικά οι τομείς του λόγου οι οποίοι επηρεάζονται ιδιαίτερα από την αυτιστική διαταραχή, τόσο στην Κατανόηση όσο και στην Έκφραση, είναι αυτοί που σχετίζονται με τα παραλεκτικά, προσωδιακά στοιχεία της ομιλίας (χροιά φωνής, ένταση, δύναμη, επιτονισμός), και με το πραγματολογικό επίπεδο του λόγου τόσο σε λεκτικό όσο και σε μη – λεκτικό επίπεδο (στάση σώματος, απόσταση, εκφράσεις προσώπου) γεγονός που καθιστά τη χρήση της γλώσσας για επικοινωνία πολύ σημαντικά επιβαρημένη. Επομένως, το κύριο χαρακτηριστικό της γλωσσικής ανάπτυξης των αυτιστικών παιδιών δεν είναι η περιορισμένη ανάπτυξη του λόγου αλλά η μη λειτουργική του χρήση.

Οι κύριες δυσκολίες των παιδιών που βρίσκονται στο φάσμα του αυτισμού παρατηρούνται στον τομέα της φωνολογίας, της σημασιολογίας, της σύνταξης και της πραγματολογίας. Πιο συγκεκριμένα σε επίπεδο φωνολογίας παρατηρείται ιδιαίτερη ποιότητα φωνής, μονότονη ομιλία, περίεργη προσωδία, ιδιαίτερος τόνος και ύψος φωνής καθώς και ηχολαλία, δηλαδή η χωρίς κατανόηση κυριολεκτική επανάληψη λέξεων ή φράσεων που έχει πει κάποιος άλλος (άμεση ηχολαλία) ή που άκουσαν σε ανύποπτο χρόνο (ετεροχρονισμένη ηχολαλία). Σε επίπεδο σημασιολογίας οι κύριες δυσκολίες είναι η μη επαρκής χρήση του

λεξιλογίου που γνωρίζουν για να ανταλλάξουν πληροφορίες και η γενική αδυναμία να κατανοήσουν ότι ό λόγος είναι ένα μέσο που μπορεί να χρησιμοποιηθεί για να πληροφορήσει και να επηρεάσει τους ανθρώπους (Παπαντωνίου, Μ. & Καμπούρογλου, Μ., 2003). Όσον αφορά τη σύνταξη παρατηρούνται δυσκολίες στην χρήση ή στον χειρισμό ορισμένων γλωσσικών τύπων (π.χ. παράλειψη μικρών γραμματικών λέξεων, δυσκολία στη χρήση των καταλήξεων των ρημάτων) επειδή δυσκολεύονται να κατανοήσουν τη σημασιολογία τους καθώς και αντιστροφή προσωπικών αντωνυμιών. Τέλος, στον τομέα της πραγματολογίας τα παιδιά δυσκολεύονται στην χρήση γλώσσας για επικοινωνιακούς σκοπούς και στην κατανόηση μεταφορικού λόγου.

6.2 Σχεδίαση Βάσης Δεδομένων

Αρχικά, σύμφωνα με τα είδη και τις κατηγορίες των σωμάτων κειμένων που έχουν αναλυθεί στο θεωρητικό μέρος, και αυτό το corpus έχει κάποια συγκεκριμένα χαρακτηριστικά. Συγκεκριμένα, πρόκειται για ένα **μονόγλωσσο**, **ειδικό** σώμα κειμένων **προφορικού λόγου**, που περιλαμβάνει την παραγωγή προφορικού λόγου παιδιών διαγνωσμένων με ΔΑΦ. Παράλληλα, το συγκεκριμένο corpus είναι **συγχρονικό** λόγω του ότι τα δεδομένα που περιλαμβάνει έχουν συλλεχθεί κατά τη διάρκεια της σχολικής χρονιάς 2017 – 2018, **ανανεώσιμο** αφού υπάρχει η δυνατότητα εμπλουτισμού του με νέα στοιχεία κρατώντας σταθερή τη σύσταση του και μη επισημειωμένο διότι δεν έχει γίνει ανάλυση λαθών.

Το σώμα κειμένων της παρούσας ερευνητικής εργασίας περιλαμβάνει τον προφορικό λόγο τριών παιδιών με Διαταραχές του Αυτιστικού Φάσματος (ΔΑΦ) που ηχογραφήθηκαν σε τρεις διαφορετικές χρήσεις του προφορικού λόγου:

1. Περιγραφή της παρακάτω εικόνας με θέμα το πάρκο.
2. Αφήγηση με χρονική σειρά με θέμα: «Περιγράφω πως πέρασα το προηγούμενο Σαββατοκύριακο;».
3. Κατευθυντικός λόγος μέσω βοηθητικών ερωτήσεων με θέμα: «Ποιο είναι το αγαπημένο σου παιχνίδι και γιατί;»

Με βάση τις οδηγίες των Bowker & Pearson (2002), το παρόν corpus περιλαμβάνει τις παρακάτω αρχές:

- **Μέγεθος:** περίπου 75 λεπτά ηχογραφημένων συνομιλιών και 9.000 λέξεις συνολικά.
- **Αριθμός κειμένων:** 9 απομαγνητοφωνημένα κείμενα από τρία διαφορετικά παιδιά.
- **Γνωστικό πεδίο:** προφορικός λόγος παιδιών με ΔΑΦ και μη τυπική ανάπτυξη της γλώσσας.
- **Περιεχόμενο:** προφορικά δεδομένα από μαγνητοφωνημένες συνομιλίες και μη προσχεδιασμένο λόγο των παιδιών
- **Δειγματοληψία:** μεταγραφές ολοκληρωμένων προφορικών δεδομένων σε τρία διαφορετικά κειμενικά είδη λόγου.
- **Αντιπροσωπευτικότητα:** 3 διαφορετικά κειμενικά είδη (περιγραφή, αφήγηση και κατευθυντικός λόγος) που έχουν παραχθεί με φυσικό τρόπο και βρίσκονται σε άμεση συνάρτηση με τους στόχους που υπηρετούνται.
- **Δημιουργός / Συγγραφέας:** κείμενα που μεταγράφηκαν από προφορικό λόγο τριών αγοριών με διάγνωση ΔΑΦ.
- **Γλώσσα:** Ελληνική.
- **Ημερομηνία έκδοσης:** σχολικό έτος 2017 – 2018.

- **Διαθεσιμότητα**: ελεύθερα διαθέσιμο στην ιστοσελίδα του εργαστηρίου Γλωσσολογίας +ΜορΦωΣη του Τμήματος Ελληνικής Φιλολογίας του Δ.Π.Θ. http://synmorphose.compulaw.gr.
- **Τήρηση μεταδεδομένων:** λεπτομερής και συστηματική κωδικοποίηση των στοιχείων που θα περιλαμβάνει το κάθε απομαγνητοφωνημένο κείμενο και η τήρηση αναλυτικών μεταδεδομένων, που θα περιλαμβάνουν πληροφορίες για τον τρόπο με τον οποίο θα γίνονται αντιληπτά και θα αναλύονται τα συστατικά μέρη του σώματος κειμένων (Γούτσος, Δ. & Φραγκάκη, Γ., 2015).

Όσον αφορά τη χρησιμότητα του προφορικού σώματος κειμένων σκοπός είναι να αποτελέσει μία σαφώς καθορισμένη συλλογή προφορικού λόγου παιδιών με ΔΑΦ στα ελληνικά και ένα εργαλείο αυτόματου εντοπισμού λαθών στον προφορικό λόγο παιδαγωγικά αξιοποιήσιμο που θα μελετήσει την γλωσσική έκφραση των παιδιών με ΔΑΦ και τα εκφραστικά μέσα που χρησιμοποιούν στον προφορικό λόγο.

7. Προδιαγραφές της εφαρμογής των σωμάτων κειμένων

Η δημιουργία του σώματος κειμένων και κατ' επέκταση της διαδικτυακής εφαρμογής θα γίνει με τη χρήση του λογισμικού WordSmith Tools 6. Δημιουργήθηκε από το Mike Scott στο Liverpool το 1996 και πρόκειται για ένα εργαλείο λεξιλογικής ανάλυσης μέσω του οποίου ο χρήστης μπορεί να μελετήσει τον τρόπο με τον οποίο οι λέξεις συμπεριφέρονται μέσα στο κείμενο κι όχι μέσα στη γλώσσα, μπορεί να αποθηκεύσει ένα μεγάλο όγκο δεδομένων με ιδιαίτερα μεγάλη ταχύτητα απόκρισης, χρησιμοποιείται για εκπαιδευτικούς σκοπούς και έχει όλα τα βασικά μεθοδολογικά εργαλεία επεξεργασίας των κειμένων που υπάγονται στο σώμα κειμένων και οδηγεί τον ερευνητή σε ασφαλή συμπεράσματα. Επίσης, η υλοποίηση της διεπαφής χρήστη θα γίνεται με τεχνολογίες Web και θα είναι φιλική προς το χρήστη (Scott, 2012, 2015).

Στη βάση δεδομένων θα υπάρχουν οι εξής τύποι αντικειμένων: α) πίνακες που θα περιέχουν όλα τα κείμενα και θα χωρίζονται σε κύριους πίνακες και βοηθητικούς, β) ευρετήρια που θα επιταχύνουν την αναζήτηση των όρων του κειμένου και γ) «αποθηκευμένες διαδικασίες και συναρτήσεις» (stored procedures, functions) δηλαδή κωδικοποιημένα ερωτήματα που θα χρησιμοποιούν οι χρήστες στην αναζήτηση.

Στην αρχική σελίδα της εφαρμογής θα εμφανίζεται η οθόνη «Καλωσορίσματος» με πληροφοριακά στοιχεία σχετικά με την εφαρμογή και το σώμα κειμένων στο οποίο βασίζεται. Ακόμα, ο χρήστης θα μπορεί να αποθηκεύει τα αρχεία κειμένου στη βάση και στη συνέχεια να προχωράει στην λεκτική ανάλυση των κειμένων. Παράλληλα, μέσω της εφαρμογής εξαγωγής λέξεων κλειδιών ο χρήστης θα παράγει για κάθε κείμενο τις λέξεις κλειδιά, ώστε να τις εμφανίσει αργότερα στη διαδικτυακή εφαρμογή σαν επιπρόσθετη πληροφορία. Επίσης, στην οθόνη της εφαρμογής με «Πληροφορίες» θα δίνονται οδηγίες για τον τρόπο αναζήτησης και ανεύρεσης προτάσεων, για το σώμα κειμένων (ημερομηνίες μαγνητοφωνήσεων, συνολικός αριθμός λέξεων), ενώ θα υπάρχει ο κατάλογος με τις μαγνητοφωνημένες συνομιλίες και τα γραπτά κείμενα των παιδιών.

Επιπλέον, ο χρήστης α) θα έχει πρόσβαση σε συμφραστικό πίνακα στον οποίο θα εμφανίζεται η λέξη ανά παιδί ή ανά κατηγορία/είδος λόγου, β) θα μπορεί να κάνει ανεύρεση λέξεων και προτάσεων σε όλο το σώμα κειμένων καθώς και σε οποιοδήποτε υποσύνολο επιλέξει, γ) θα έχει τη δυνατότητα να ανατρέξει σε υποσύνολα κειμένων που έχει ορίσει και αποθηκεύσει στο παρελθόν και δ) θα

καθορίζει το μέγιστο αριθμό προτάσεων ανά σελίδα αποτελεσμάτων αλλά και πόσες προτάσεις θα του επιστραφούν από το σύστημα.

Όσον αφορά την αναζήτηση, απλή ή σύνθετη, δίνονται οι παρακάτω διευκρινίσεις:

- Κάθε συμφραστικός πίνακας θα δίνει 3 στήλες (η μεσαία θα περιέχει τη λέξη που αναζητήσαμε).
- Ο χρήστης θα μπορεί να κάνει αναζήτηση ανά είδος κειμένου (περιγραφικό, αφηγηματικό ή κατευθυντικό) ή ανά παιδί.
- Η αναζήτηση θα μπορεί να γίνει με βάση μία λέξη, ένα λήμμα ή γραμματικές εκφράσεις (ή συνδυασμό των κριτηρίων αυτών).
- Η λέξη η οποία θα πληκτρολογείται στην αναζήτηση δεν θα έχει σημασία αν θα έχει τόνο ή όχι, δηλαδή η εξεύρεση της λέξης *παιδί* επιτυγχάνεται και με την άτονη λέξη *παιδι*.
- Για αντικατάσταση ενός ή περισσοτέρων γραμμάτων στη λέξη αναζήτησης θα χρησιμοποιούνται κάποιοι ειδικοί χαρακτήρες στη θέση κάποιου γράμματος μέσα στον όρο που αναζητείται.
- Η αναζήτηση για μία λέξη θα δίνει αυτόματα όλες τις λέξεις που αρχίζουν με την ακολουθία της λέξης που αναζητήσαμε.

  Π.χ. στην αναζήτηση της λέξης *τσουλήθρα* θα εμφανιστούν λέξεις με την ίδια ακολουθία γραμμάτων με τη λέξη που αναζητήσαμε, όπως π.χ. *τσουλήρθα, τσουλήθα ή τσουλήρθες*.
- Χρησιμοποιώντας τον αστερίσκο (*) στην αναζήτηση μίας λέξης θα εμφανίζονται αυτόματα όλες οι λέξεις που περιέχουν την ακολουθία της λέξης ή λήγουν σε αυτή.

  Π.χ. αν η λέξη που αναζητήσουμε είναι **θέα** ή **θέα* θα εμφανίζονται πλην της λέξης *θέα* και όλες οι λέξεις που περιέχουν την ακολουθία θ-ε-α ή λήγουν σε αυτήν, π.χ. *αξιοθέατα, άθεα*.
- Η λέξη που αναζητούμε δεν θα έχει σημασία αν είναι γραμμένη με κεφαλαία ή πεζά γράμματα.

  Π.χ. η εξεύρεση της λέξης *ξύπνησα* επιτυγχάνεται και με τη λέξη *ΞΥΠΝΗΣΑ*.
- Σε αναζήτηση που εμφανίζονται σύμβολα απομαγνητοφώνησης θα δίνεται η σημασία τους.
- Στη λέξη που θα αναζητήσουμε με ένα ¨κλικ¨ πάνω της θα εμφανίζεται η προηγούμενη και η επόμενη λέξη κατά σειρά.

  Π.χ. στην αναζήτηση της λέξης *'πράσινο'* θα εμφανίζεται η ακολουθία των λέξεων: *πράσινο - πράσινο καπέλο* (επανάληψη λέξης).
- Με δύο ¨κλικ¨ πάνω στη λέξη θα εμφανίζεται η λέξη μέσα στην πρόταση στην οποία περιλαμβάνεται.

  Π.χ. στη λέξη *'πράσινο'* που αναζητήθηκε παραπάνω θα εμφανιστεί με δύο "κλικ" η πρόταση: *Ο άντρας φοράει πράσινο – πράσινο καπέλο.*

8. Συμπεράσματα-Περιορισμοί της έρευνας

Όσον αφορά το γραπτό σώμα κειμένων παρατηρήθηκε απουσία λαθών στα κείμενα που ανήκουν στην κατηγορία της αντιγραφής. Επίσης, τα περισσότερα ορθογραφικά λάθη εντοπίστηκαν στη γραφή καθ' υπαγόρευση λόγω προβλημάτων φωνημικής, μορφολογικής και οπτικοορθογραφικής επίγνωσης αλλά και ακουστικής μνήμης και διάκρισης, ενώ υπήρξαν λίγα λάθη στην κατηγορία ελεύθερης παραγωγής γραπτού λόγου, καθώς οι μαθητές χρησιμοποιούν λέξεις που γνωρίζουν τον τρόπο γραφής τους. Επιπλέον, στην

κατηγορία της αντιγραφής η ποιότητα των γραμμάτων ήταν καλύτερη από τα κείμενα καθ' υπαγόρευσης. Τα λάθη μεγαλύτερης συχνότητας ήταν λάθη παραλείψεων ή αντικαταστάσεων φωνημάτων στην κατηγορία της φωνολογίας, λάθη στις καταλήξεις των ρημάτων και των ουσιαστικών στην κατηγορία της γραμματικής, λάθη στο θέμα των λέξεων και κυρίως των πιο σπάνιων, απουσία τόνων και δυσκολία στη χρήση κεφαλαίου.

Σχετικά με το προφορικό σώμα κειμένων τα λάθη που παρατηρήθηκαν είναι φωνολογικά, σημασιολογικά, συντακτικά και πραγματολογικά. Λάθη μεγαλύτερης συχνότητας ήταν στον τομέα της πραγματολογίας.

Συμπερασματικά, τα corpora που δημιουργήθηκαν στην παρούσα εργασία αποτελούν δύο βασικά εργαλεία έρευνας για τα ελληνικά δεδομένα της ειδικής εκπαίδευσης μέσα από τα οποία σε μεταγενέστερο στάδιο ο χρήστης θα εντοπίζει τα λάθη των παιδιών με ΕΜΔ – δυσορθογραφία στο γραπτό λόγο καθώς και τα λάθη των παιδών με ΔΑΦ στον προφορικό λόγο. Επομένως, ο χρήστης έχει τη δυνατότητα να μελετήσει και να συγκρίνει τα λάθη προς όφελος δικό του αλλά και των μαθητών του.

Στον αντίποδα, στα πλαίσια της ερευνητικής μας εργασίας υπάρχουν και πτυχές οι οποίες δεν αναπτύχθηκαν όπως θα έπρεπε με αποτέλεσμα η έρευνά μας να περιορίζεται σε κάποιους τομείς. Αρχικά, δεν έγινε εξαντλητική έρευνα και το δείγμα μας δεν ήταν επαρκές, καθώς θα μπορούσε να περιλαμβάνει περισσότερα παιδιά ως δείγματα από διάφορες εκπαιδευτικές βαθμίδες. Επίσης, είναι χρήσιμη η συμμετοχή κοριτσιών ως δείγμα για τη σύγκριση ανάμεσα στα δύο φύλα. Επιπλέον, χρήσιμο θα ήταν να συμπεριληφθούν στην έρευνα περισσότερα κειμενικά και προφορικά είδη και επιστημονικά πεδία. Ακόμα, ένας ακόμα περιορισμός στον οποίο υπόκειται η εργασία μας είναι ότι δεν προχωρήσαμε στην υλοποίηση της εφαρμογής βάσης δεδομένων στην οποία θα περιέχονται όλα τα δείγματα των απομαγνητοφωνημένων κειμένων και θα παρέχονται με ελεύθερη πρόσβαση στο διαδίκτυο. Τέλος, ενώ η επισημείωση είναι χρήσιμη, μπορεί να δυσκολέψει το χρήστη κάνοντας περίπλοκη τη χρήση του corpus.

Συνολικά, πρέπει να σημειωθεί ότι τα σώματα κειμένων που δημιουργήσαμε είναι δυναμικά και ανοικτά γεγονός που αποτελεί έναυσμα για περαιτέρω επεξεργασία με σκοπό τον εμπλουτισμό, τη διόρθωση και την ανανέωσή τους.

9. Βιβλιογραφία

Alamri, M., & Teahan, W. J. (2017). A New Error Annotation for Dyslexic texts in Arabic. In The Association for Computational Linguistics (Ed.). *The Third Arabic Natural Language Processing Workshop. Proceedings of the Workshop* (pp. 72-78). Spain: The Association for Computational Linguistics.

American Psychiatric Asociation. (2000). *Diagnostic and statistical manual of mental disorders (4th 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.

Barnard D. & Ide N. (1995). *The Text Encoding Initiative: Flexible and Extensible Document Encoding*, Canada: Queen's University, Kingston.

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.

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. (2003). *Σώμα Ελληνικών Κειμένων: Σχεδιασμός και υλοποίηση. Πρακτικά του 6ου Διεθνούς Συνεδρίου Ελληνικής Γλωσσολογίας*, Κρήτη: Πανεπιστήμιο Κρήτης. Διαθέσιμο στο: http://www.philology.uoc.gr/conferences/6thICGL/gr.htm [Πρόσβαση 4 Ιανουαρίου 2018].

Γούτσος, Δ. &Φραγκάκη, Γ. (2014). Πρόσφατη γλωσσική αλλαγή στα ελληνικά: Σχεδιασμός του Διαχρονικού Σώματος Ελληνικών Κειμένων του 20ού αιώνα. Στο G. Kotzoglou, K. Nikolou, E. Karantzola, K. Frantzi, I. Galantomos, M. Georgalidou, V. Kourti-Kazoullis, C. Papadopoulou&E. Vlachou (Επιμ.). *Selected Papers of the 11th International Conference on Greek Linguistics* (σελ. 318-329). Rhodes: University of the Aegean. Διαθέσιμοστο: http://www.rhodes.aegean.gr/Assets/announcements/diafora/ICGL_11_selected_papers.pdf [Πρόσβαση 4 Ιανουαρίου 2018].

Γούτσος, Δ., &Φραγκάκη, Γ. (2015). *Εισαγωγή στη γλωσσολογία σωμάτων κειμένων*. Αθήνα: ΣΕΑΒ.

Gavriilidou, Z., & Kambaki-Vougioukli, P. (2011). Phonological Description of the Greek Speaking Children Corpus (GSCC). In K. Chatzopoulou, A. Ioannidou, & Y. Suwon (Ed.). *Proceedings of the 9th ICGL* (pp. 559-568). Available at http://www.ling.ohiostate.edu/ICGL/proceedings/48_Gavriilidou_Kambaki_ED_559.pdf [Accessed 14 April 2018].

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Kennedy, G. (1998). *An introduction to corpus linguistics*. London and New York: Longman.

Leech, G. (1997). Grammatical tagging. In R. Garside, G. Leech, & T. Mc Enery (Ed.). *Corpus Annotation* (pp. 19-33). London: Routledge.

Mair, C., Hundt, M., Leech, G. N., & Smith, N. (2002). Short term diachronic shifts in part-of-speech frequencies: a comparison of the tagged LOB and F-LOB corpora. *International Journal of Corpus Linguistics*, 7(2), 245-264.

McEnery, T. & Wilson, A. (1996). *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

McEnery, T. & Xiao, R. (2007). Parallel and comparable corpora. What is happening? G. James, & G. Anderman (Ed.). *Incorporating Corpora: The Linguist and theTranslator* (pp.18-31). Clevedon, UK: Multilingual Matters.

Μικρός, Γ. (2004). Ηλεκτρονικά σώματα κειμένων και ορολογία. Στο Μ. Κατσογιάννου &Ε. Ευθυμίου (Επιμ.). *Ελληνική ορολογία: Έρευνα και εφαρμογές* (σελ. 139-164). Αθήνα: Καστανιώτης.

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 [Τελευταία Πρόσβαση: 5 Μαρτίου 2018].

Πόρποδας, Κ. Δ. (2002). *Η ανάγνωση*. Πάτρα: χ.ό.

Rello, L. (2014). Design of word exercises for children with dyslexia. *Procedia Computer Science*, *27*, 74-83.

Rollins, P. (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.

Scott, M. (2012). *WordSmith Tools*. Available at: https://www.leuphana.de/fileadmin/user_upload/PERSONALPAGES/_abcd/barron_anne/files/WordSmith6.pdf [Accessed 17 May 2018].

Scott, M. (2015). *Wordsmith Tools Help*. Liverpool: Lexical Analysis Software.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). *Preliminary recommendations on corpus typology. EAGLES document*. Available at:http://www.ilc.pi.cnr.it/EAGLES/corpustyp/corpustyp.html [Accessed 4 January 2018].

Sinclair, J. (2005). Corpus and Text. Basic Principles. In M. Wynne (Ed.). *Developing Linguistic Corpora: A Guide to Good Practice* (pp. 1-16). Oxford: Oxbow Books, Available at:http://www.ahds.ac.uk/creating/guides/linguisticcorpora/chapter1.htm [Accessed 4 January 2018].

Sperberg-McQueen, C.M. and Burnard, L. (1994). *Guidelines for electronic text encoding and interchange*, TEI-P3, Chicago and Oxford: ACH-ACL-ALLC Text Coding Initiative.

Statkiewicz-Gayhard, V., Peerenboom, B. & Campell, R.N. (2001). Διασχίζοντας τις γέφυρες. Η γονεική προοπτική στην αντιμετώπιση ενός παιδιού αφού έχει διαγνωστεί με Αυτισμό/Διάχυτη αναπτυξιακή Διαταραχή. (Α. Λαζαρίνη, Μεταφρ.) Αθήνα: ΕΕΠΑΑ.

Štefanec, V., Ljubešić, N., &Kraljević, J. K. (2016, January). Croatian Error-Annotated Corpus of Non-Professional Written Language. In *LREC 2016, Tenth International Conference on Language Resources and Evaluation*.

World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: World Health Organization.