



# EURALEX XIX

Congress of the  
European Association  
for Lexicography

Lexicography for inclusion

**7-11 September 2021**

Ramada Plaza Thraki

Alexandroupolis, Greece

[www.euralex2020.gr](http://www.euralex2020.gr)

**Proceedings Book  
Volume 1**

Edited by Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

---

**EURALEX Proceedings**

ISSN 2521-7100

ISBN 978-618-85138-1-5

Edited by: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

**2020 Edition**

---

# The Greek Children Spoken Language Corpus (GCSL Corpus) / Το Προφορικό Σώμα Κειμένων Ελληνόφωνων Παιδιών (ΠΣΚΕΠ): Παρουσίαση, Εφαρμογές και Προοπτικές

Motsiou E.

University of Thessaly, Greece

## Περίληψη

Με την εξέλιξη της τεχνολογίας και τη δυνατότητα δημιουργίας σωμάτων κειμένων που περιέχουν μεγάλο όγκο δεδομένων η μελέτη της γλώσσας διευκολύνθηκε σημαντικά. Ειδικά όσον αφορά την έρευνα του αυθόρμητου λόγου, τα προφορικά σώματα κειμένων προσφέρουν μια ποικιλία και αριθμό δεδομένων που θα ήταν δύσκολο να αποκτηθεί με κάποια άλλη μέθοδο συλλογής. Το Προφορικό Σώμα Κειμένων Ελληνόφωνων Παιδιών (ΠΣΚΕΠ/ GCSL Corpus) έχει σαν σκοπό να διευκολύνει τη μελέτη για τη γλωσσική ανάπτυξη δίνοντας σε φοιτητές και ερευνητές ελεύθερη πρόσβαση σε υλικό για την εξελισσόμενη γλώσσα παιδιών προσχολικής ηλικίας των οποίων πρώτη γλώσσα είναι η Ελληνική. Το Σώμα αποτελείται από ψηφιακά αρχεία με συνομιλίες ενηλίκων και παιδιών, τα οποία εμπλουτίζονται σταδιακά, και προσφέρει τη δυνατότητα στο χρήστη να επιλέξει συνομιλίες χρησιμοποιώντας παράλληλα έναν συνδυασμό κριτηρίων αναζήτησης.

**Λέξεις-κλειδιά:** προσχολική ηλικία, αυθόρμητος λόγος, προφορικά σώματα κειμένων

## 1 Εισαγωγή: Η Μελέτη του Παιδικού Λόγου και τα Σώματα Κειμένων

Η μελέτη της εξέλιξης της παιδικής γλώσσας και η περιγραφή των ιδιαίτερων χαρακτηριστικών της είναι ένα γνωστικό πεδίο που έχει απασχολήσει ερευνητές διάφορων επιστημονικών κλάδων συναφών με την ανάπτυξη του παιδιού. Μια τέτοιο είδους έρευνα απαιτεί την καταγραφή και την ανάλυση μεγάλου όγκου δεδομένων από την παιδική γλώσσα. Με την ανάπτυξη της τεχνολογίας, η συλλογή υλικού από την παρατήρηση/ καταγραφή του αυθόρμητου παιδικού λόγου και από τις πειραματικές μεθόδους διευκολύνθηκε σημαντικά. Σήμερα, η εκτεταμένη αξιοποίηση των τεχνολογιών στην ανάλυση της γλώσσας και τη δημιουργία βάσεων δεδομένων έχει εξασφαλίσει τη δυνατότητα καταγραφής ενός εκτεταμένου και ποικίλου υλικού από τον παιδικό λόγο, την κωδικοποίησή του και την αυτόματη ανάλυση μεταβλητών, ώστε η ανάπτυξη της επιστήμης της παιδικής γλώσσας να διαθέτει τεράστιες προοπτικές για το μέλλον.

Παρότι ο αυθόρμητος λόγος είναι κεντρικό κομμάτι για τη μελέτη της γλώσσας και αναγνωρίζεται διεθνώς ότι τα προφορικά σώματα κειμένων είναι απαραίτητο εργαλείο για την έρευνα, τα δεδομένα από τον προφορικό λόγο είναι γενικά ελλιπή, εν μέρει εξαιτίας των δυσκολιών που έχουν να κάνουν με τη συλλογή τους (Pavlidou 2012). Το ίδιο ισχύει και για τον παιδικό λόγο, για τον οποίο ο αριθμός των βάσεων δεδομένων αλλά και η προσβασιμότητα είναι αρκετά περιορισμένα. Η πιο γνωστή διεθνής βάση δεδομένων παιδικού λόγου αναπτύχθηκε στην Αμερική το 1984 και ονομάστηκε CHILDES (MacWhinney 2000). Η βάση αυτή περιέχει δεδομένα και από την ελληνική γλώσσα, ωστόσο είναι μικρής σχετικά έκτασης.

Είναι φανερό ότι μια συλλογή γλωσσικού υλικού από παιδιά που καταγράφει σύγχρονα και διαρκώς αυξανόμενα δεδομένα, επιτρέποντας την ανάλυση και την εξαγωγή συμπερασμάτων για την αναπτυσσόμενη γλώσσα διευκολύνει και προωθεί τόσο τη μελέτη της φυσιολογικής ανάπτυξης της παιδικής γλώσσας όσο και της αποκλίνουσας, αλλά και τροφοδοτεί με κρίσιμα πορίσματα το χώρο της εκπαίδευσης και διδακτικής της γλώσσας (Behrens 2008; Diessel 2009). Το Προφορικό Σώμα Κειμένων Ελληνόφωνων Παιδιών (ΠΣΚΕΠ) έχει σαν σκοπό να διευκολύνει την έρευνα για τη γλωσσική ανάπτυξη δίνοντας σε φοιτητές και ερευνητές πρόσβαση σε υλικό για τη γλώσσα μικρών παιδιών με μητρική την Ελληνική. Το υλικό είναι μια συλλογή συνομιλιών ενηλίκων και παιδιών προσχολικής ηλικίας, συνεντός σε ένα μεγάλο μέρος αποτελεί πηγή αυθόρμητου παιδικού λόγου.

## 2 Το Προφορικό Σώμα Κειμένων Ελληνόφωνων Παιδιών (ΠΣΚΕΠ)

Η ανάπτυξη ενός προφορικού σώματος κειμένων τυπικά ακολουθεί τις εξής διαδικασίες: διερεύνηση/ επέκταση πηγών, μεταγραφή, καθαρισμός-αποθήκευση, τυποποίηση, κωδικοποίηση και επισήμειωση (Γούτσος & Φραγκάκη 2015, Wynne 2005 κ.ά.). Η πηγές, οι διαδικασίες συλλογής και μεταγραφής των δεδομένων, καθώς και τα βασικά χαρακτηριστικά του ΠΣΚΕΠ περιγράφονται αμέσως παρακάτω.

### 2.1 Τα Δεδομένα

Το υλικό του ΠΣΚΕΠ αποτελείται από αυθεντικές συνομιλίες ενηλίκων και παιδιών προσχολικής ηλικίας. Ο συνολικός αριθμός των παιδιών, των οποίων ο λόγος καταγράφεται, είναι μέχρι σήμερα 91, 49 κορίτσια και 42 αγόρια, με ΜΟ ηλικίας τα 4 έτη. Όλα τα παιδιά είναι τυπικά αναπτυσσόμενα, μονόγλωσσα ή δίγλωσσα, που ζουν σε αστικές, ημιαστικές ή αγροτικές περιοχές (για παράδειγμα, Βόλος, Καβάλα, Αθήνα, Αγία Λάρισα, Αλόνησος, Νεοχώρι Καρδίτσας κτλ.).

## 2.2 Διαδικασία Συλλογής Υλικού

Το υλικό συγκεντρώνεται, από το 2015 και εξής, με τη βοήθεια των φοιτητών του Παιδαγωγικού Τμήματος Προσχολικής Εκπαίδευσης του Πανεπιστημίου Θεσσαλίας, στα πλαίσια του ετήσιου μαθήματος «Ανάπτυξη του λόγου στο παιδί». Ύστερα από κατάλληλη εκπαίδευση, οι φοιτητές/ φοιτήτριες μαγνητοφωνούν συνομιλίες με παιδιά (δικές τους ή άλλων), έχοντας εξασφαλίσει τη συγκατάθεση των γονέων και έχοντας πάρει όλα τα απαραίτητα μέτρα προστασίας προσωπικών δεδομένων. Από τους φοιτητές πραγματοποιείται και μια πρώτη απομαγνητοφώνηση του υλικού.

## 2.3 Μεταγραφή Δεδομένων και Προετοιμασία της Βάσης του ΠΣΚΕΠ

Η κύρια φάση της μεταγραφής των μαγνητοφωνημένων δεδομένων συνεχίστηκε μετά από τη σύσταση ομάδας έργου<sup>1</sup> με κατάλληλα εκπαιδευμένους συνεργάτες, οι οποίοι έχουν αναλάβει τον έλεγχο και τις διορθώσεις της πρώτης καταγραφής, την επιμέλεια και την προετοιμασία του υλικού για την εισαγωγή του στη βάση. Ακολουθεί μια τρίτη φάση αντιπαραβολής, ελέγχου και διόρθωσης από την κύρια ερευνήτρια. Κάθε συνομιλία αποθηκεύεται σε μορφή word doc (εικόνα 1) μαζί με τα μεταδεδομένα για την «ταυτότητα» κάθε συνομιλίας (πίνακας 1).

<b>E:</b> Δε σου αρέσει η τσιχλόφουσα; <b>E;</b>
<b>N:</b> οοο
<b>E:</b> Δε σ' αρέσει; Θυμάσαι που κάναμε μπάνιο στη θάλασσα;
<b>N:</b> ne
<b>E:</b> Εε, φορές μπρατσάκια; Για πες μου, εδώ φορές μπρατσάκια; Ή ξέρεις να κολυμπάς μόνη σου;
<b>N:</b> pa'l'a xron'a ee... evaza bratsak'a ala tora δε vazο
<b>E:</b> Βατραχοπέδιλα βάζεις;
<b>N:</b> ο ((αρνητική απάντηση)) ee... m bori k'e ne
<b>E:</b> Μπορεί και να; Α, δηλαδή ξέρεις να κολυμπάς και λίγο ε; Κάνεις βουτιές;
<b>N:</b> ne
<b>E:</b> Πολύ μεγάλες κάνεις βαθιά;
<b>N:</b> na ο nonos mu me petai sti thalasa TOSO psila ((δείχνει πόσο ψηλά την πετάει ο νονός της))

Εικόνα 1: Απομαγνητοφωνημένη & μεταγεγραμμένη συνομιλία

<b>Αριθμός συνομιλίας:</b> (σειρά συνομιλίας στο Σώμα)
<b>Συλλογή υλικού:</b> Κωδικός ερευνητή/τριας που συλλέγει το υλικό
<b>Απομαγνητοφώνηση:</b> κωδικός ερευνητή/τριας που απομαγνητοφωνεί το υλικό
<b>Συμμετέχοντες:</b> αναφορά συμμετεχόντων στη συνομιλία
<b>Φύλο νηπίου:</b> κορίτσι/ αγόρι
<b>Ηλικία νηπίου:</b> έτη και μήνες
<b>Χώρος:</b> χώρος διεξαγωγής συνομιλίας, π.χ. σπίτι, νηπιαγωγείο κτλ
<b>Τόπος διαμονής:</b> γεωγραφικός χώρος διαμονής νηπίου
<b>Πλαίσιο:</b> περιγραφή πλαισίου επικοινωνίας κατά τη διάρκεια της συνομιλίας, π.χ. παιχνίδι, αφήγηση κτλ
<b>Ημερομηνία καταγραφής:</b> Η/Μ/Ε
<b>Διάρκεια:</b> χρονική διάρκεια συνομιλίας (ΜΟ 14'30'')
<b>Κωδικός:</b> κωδικός καταχώρησης συνομιλίας (βλ.εικόνα 2)
<b>Σύμβολα συμμετεχόντων:</b> Ε (ερευνητής/τρια), Ν (νήπιο), Μ (μητέρα), Π (πατέρας), Σ (συμμετέχων/ουσα: φιλικά ή λοιπά συγγενικά πρόσωπα νηπίου ή ερευνητών ή γονέων)

Πίνακας 1: Καρτέλα στοιχείων συνομιλιών (μεταδεδομένα)

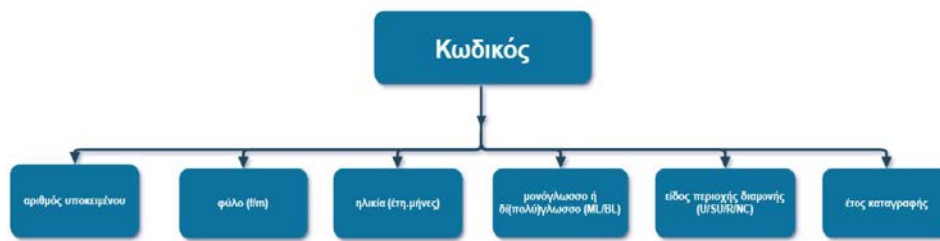
<sup>1</sup> Η ομάδα έργου που δημιούργησε το ΠΣΚΕΠ αποτελείται από τους:

Ελένη Μότσιου, Επίκουρη καθηγήτρια Παιδαγωγικού Τμήματος Προσχολικής Εκπαίδευσης του Πανεπιστημίου Θεσσαλίας, επιστημονικά υπεύθυνη

Θάνο Λίτσο, Μεταπτυχιακό φοιτητή του Τμήματος Αγγλικής Γλώσσας και Φιλολογίας του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, υπεύθυνο για την μεταγραφή και επιμέλεια των αρχείων του Σώματος

Γιάννη Κυριαζή, Ηλεκτρονικό Μηχανικό ΤΕ στο Παιδαγωγικό Τμήμα Προσχολικής Εκπαίδευσης του Πανεπιστημίου Θεσσαλίας, υπεύθυνο για τη δημιουργία της ηλεκτρονικής βάσης δεδομένων και ιστοσελίδας του ΠΣΚΕΠ.

Ο κωδικός κάθε συνομιλίας αποτελείται από τα εξής στοιχεία:



Εικόνα 2: Ανάλυση κωδικού συνομιλιών

Ο κωδικός εξειδικεύει επιπλέον ανάμεσα σε μονόγλωσσο ή δίγλωσσο παιδί (ML/ BL) και είδος γεωγραφικής περιοχής διαμονής του νηπίου (U (urban)/ SU (semi-urban)/ R (rural)/ NC(non categorized), στοιχεία που μπορούν να χρησιμοποιηθούν και ως κριτήρια επιλογής στο κουμπί Αναζήτηση.

Ο λόγος του ερευνητή/-τριας ή άλλων συμμετεχόντων καταγράφεται ορθογραφικά. Κατά την κωδικοποίηση του παιδικού λόγου επιδίωξη ήταν να αποδοθούν και φωνητικές ιδιαιτερότητες/ αποκλίσεις. Έτσι, για την καταγραφή του, χρησιμοποιήθηκε κατά βάση το Διεθνές Φωνητικό Αλφάβητο, με ορισμένες προσαρμογές για διευκόλυνση της ανάγνωσης, στην κατηγορία των ουρανικών: συγκεκριμένα, τα ουρανικά συμβολίζονται με [k', g', x', n', l'] (όμως κρατείται το σύμβολο [j] για το ουρανικό τριβόμενο ηχηρό).

Κατά την απομαγνητοφώνηση και μεταγραφή χρησιμοποιήθηκαν οι ελάχιστες δυνατές συμβάσεις/ σύμβολα, προκειμένου να διασφαλιστεί η κατά το δυνατόν ευχερής ανάγνωση της συνομιλίας και να περιοριστούν οι μη γλωσσικές πληροφορίες (Sinclair 1991). Τα σύμβολα που χρησιμοποιούνται είναι τα εξής:

- Οι παρατηρήσεις του ερευνητή που επεξηγεί/ σχολιάζει την περίπτωση ή το λόγο του παιδιού μπαίνουν σε διπλή παρένθεση: (( ))
- Εάν κάποιος φθόγγος, λέξη, φράση κτλ δεν ακούγεται καλά τοποθετείται σε μονή παρένθεση/ μια άδεια παρένθεση στη θέση του ακατανόητου εκφωνήματος: ( )
- Κεφαλαία: δηλώνουν αύξηση της έντασης της φωνής
- Αποσιωπητικά: δηλώνουν παύσεις στο λόγο
- Διπλά σύμβολα: δηλώνουν διάρκεια στην εκφώνηση

Το ΠΣΚΕΠ είναι ένα απλό, ειδικό και ανοιχτό σώμα κειμένων, που περιέχει προς το παρόν 84 συνομιλίες (128.100 λέξεις). Ως Σύστημα Διαχείρισης Περιεχομένου (CMS) έχει χρησιμοποιηθεί το ανοικτό και ελεύθερο λογισμικό WordPress. Η ιστοσελίδα έχει εγκατασταθεί σε μία πλατφόρμα LAMP (Linux/Apache/MySQL/PHP) στο χώρο του Πανεπιστημίου Θεσσαλίας.

## 2.4 Εμφάνιση, Δομή και Χαρακτηριστικά του ΠΣΚΕΠ

Πληκτρολογώντας στον φυλλομετρητή τη διεύθυνση <http://gcslece.uth.gr> ο χρήστης βρίσκεται στην αρχική σελίδα του ΠΣΚΕΠ:

Εικόνα 3: Το περιβάλλον του ΠΣΚΕΠ

Στο περιβάλλον αυτό ο χρήστης έχει τη δυνατότητα της επιλογής των παρακάτω κουμπιών:

- i. *To Corpus*: περιέχει πληροφορίες για το Σώμα και την ερευνητική ομάδα
- ii. *Περιήγηση*: ο χρήστης μπορεί να δει σε έναν συγκεντρωτικό πίνακα όλες τις καταγεγραμμένες συνομιλίες, και να επιλέξει να δει ορισμένες πατώντας τον κωδικό κάθε μιας:

Α/Α	ΣΥΛΛΟΓΗ ΥΛΙΚΟΥ	ΑΠΟΜΑΓΗΤΟΦΩΝΗΣΗ ΥΛΙΚΟΥ	ΣΥΜΜΕΤΕΧΟΝΤΕΣ	ΦΥΛΟ ΝΗΠΙΟΥ 1	ΗΛΙΚΙΑ ΝΗΠΙΟΥ 1	ΦΥΛΟ ΝΗΠΙΟΥ 2	ΗΛΙΚΙΑ ΝΗΠΙΟΥ 2	ΧΩΡΟΣ	ΤΟΠΟΣ ΔΙΑΜΟΝΗΣ	ΠΛΑΙΣΙΟ	ΗΜΕΡΟΜΗΝΙΑ	ΔΙΑΡΚΕΙΑ	ΚΩΔΙΚΟΣ 1	ΚΩΔΙΚΟΣ 2
1	Ερευνητρια Ευ. Σκ.	Ερευνητρια Ευ. Σκ.	Νήπιο, Ερευνητρια	Κορίτσι	3 ετών και 6 μηνών			Σπίτι νηπίου, αυτοκίνητο	Βόλος	Ελεύθερη παρατήρηση, συζήτηση	December 17, 2014	10' 56"	1f3.6MLU2014	
2	Ερευνητρια Μα. Μπ.	Ερευνητρια Μα. Μπ.	Νήπιο, Αδερφή νηπίου, Ερευνητρια	Αγόρι	5 ετών και 5 μηνών			Σπίτι νηπίου	Βόλος	Ελεύθερη συζήτηση	January 11, 2015	5' 02"	2m5.5MLU2015	
3	Ερευνητρια Βα. Γε.	Ερευνητρια Βα. Γε.	Νήπιο, Ερευνητρια	Αγόρι	4 ετών και 6 μηνών			Σπίτι νηπίου	Αγιά Λάρισα	Συζήτηση	April 13, 2015	10' 35"	3m4.6MLR2015	
4	Ερευνητρια Αι. Κα.	Ερευνητρια Αι. Κα.	Νήπιο, Ερευνητρια	Αγόρι	5 ετών και 3 μηνών			Σπίτι νηπίου	Καβάλα	Παιχνίδι	April 17, 2015	10' 52"	4m5.3MLU2015	
5	Ερευνητρια Μα. Πα.	Ερευνητρια Μα. Πα.	Νήπιο, Ερευνητρια	Κορίτσι	5 ετών και 2 μηνών			Σπίτι νηπίου, στο παιδικό δωμάτιο	Σέρρες	Παιχνίδι, συζήτηση	January 3, 2015	4' 55"	5f5.2MLU2015	
6	Ερευνητρια 1 Ασ. Κο.	Ερευνητρια 2 Αβ. Στ.	Νήπιο, Ερευνητρια	Κορίτσι	4 ετών και 2 μηνών			Σπίτι νηπίου	-	Συζήτηση	December 1, 2006	-	6f4.2MLNC2016	
7	Ερευνητρια Γκ. Θω.	Ερευνητρια Γκ. Θω.	Νήπιο, Ερευνητρια	Κορίτσι	4 ετών και 5 μηνών			Σπίτι νηπίου	Λάρισα	Συζήτηση και άνεληση πληροφοριών μέσα από το παιχνίδι (ζωγραφική, επιτραπέζια παιχνίδια και...	June 11, 2015	1705"	7f4.5MLU2015	

Εικόνα 4: Περιήγηση – συγκεντρωτικός πίνακας συνομιλιών

Επιλέγοντας έναν από τους κωδικούς ο χρήστης μπορεί να δει ολόκληρη τη συνομιλία που αντιστοιχεί σε αυτόν:

Αριθμός συνομιλίας:	15	<b>Ε:</b> Ποια είναι τα αγαπημένα σου παραμύθια;
Συλλογή υλικού:	Ερευνητρια 1 Πν. Να.	<b>Ν:</b>
Απομαγνητοφώνηση:	Ερευνητρια 2 Σο. Ηλ.	<b>Ε:</b> Απ' όλα ποιο σ'αρέσει πιο πολύ;
Συμμετέχοντες:	Νήπιο, Ερευνητρια 1	<b>Ν:</b> mmm apo afta ru exo spiti mu?
Φύλο νηπίου:	Κορίτσι	<b>Ε:</b> Από αυτά που έχεις σπίτι σου, από αυτά που έχετε εδώ, ποιο σ'αρέσει πιο πολύ;
Ηλικία νηπίου:	5 ετών	<b>Ν:</b> a...., a...., afta ru maresun px'o poli ine
Χώρος:	Τάξη νηπιαγωγείου	<b>Ε:</b> Φέρτο μαζί σου.
Τόπος διαμονής:	Βόλος	<b>Ν:</b> Όλο mu ine.
Πλαίσιο:	Κατευθυνόμενη συζήτηση	<b>Ε:</b> Δικό σου είναι;
Ημερομηνία καταγραφής:	30 Ιουν. 1905	<b>Ν:</b> ne.
Διάρκεια:	10'	<b>Ε:</b> Για έλα εδώ να κάτσουμε.
Κωδικός:	15f5MLU2008	<b>Ν:</b> ojanase to na dis.
Σύμβολα συμμετεχόντων:	N, E	<b>Ε:</b> Ξέρεις τι, Δε θα το διαβάσω εγώ, θα μου πεις εσύ τι λέει. Θέλεις, Να μου πεις περίπου τι λέει, Πώς λέγεται;
		<b>Ν:</b> ammmmm.... I anak'iklosi.
		<b>Ε:</b> Και η φίλη της.
		<b>Ν:</b> ne k'e i anak'iklosi k'e i fili tis.
		<b>Ε:</b> Θέλεις, όταν το ανοίγουμε, να μου δείχνεις τι λέει.
		<b>Ν:</b> ox'i, den ksero na ojanazo.
		<b>Ε:</b> Φυσικά δεν ξέρεις να διαβάζεις, αλλά να μου δείχνεις περίπου την ιστορία. Τι δείχνει εδώ; Γιατί σ'αρέσει αυτό το παραμύθι;
		<b>Ν:</b> jati....., jati aftos o dukalakis e illi, tha... se liyo eridi adjasan tha ksana jinun opos prin.
		<b>Ε:</b> Αλήθεια; Θα γίνουν όπως πριν; Για δείξε μου πού γίνονται όπως πριν;
		<b>Ν:</b> ru?
		<b>Ε:</b> Πού γίνονται όπως πριν. Πιο κοντά;
		<b>Ν:</b> ne, ala na sas po kali?
		<b>Ε:</b> Ναι ό,τι θέλεις.
		<b>Ν:</b> eeee, afto ix'e kolon'a mesa.
		<b>Ε:</b> Α, είχε κολόνια, εεε; Αυτό τι είχε μέσα;
		<b>Ν:</b> k'e adjase afto k'e to pefaksa k'e afto ix'e. ili itane vazο me luluδja xoris tet'ο.
		<b>Ε:</b> Αααα, και αυτό τώρα τι είναι;

Εικόνα 5: Περιβάλλον συνομιλιών

*Σημείωση:* μπορεί να γίνει αναζήτηση συγκεκριμένων λέξεων στην επιλεγμένη σελίδα μέσω φυλλομετρητή (browser), πατώντας Ctrl+F στο πληκτρολόγιο και εισάγοντας το γλωσσικό στοιχείο: η λέξη "μαρκάρεται" ανάλογα και με τα κριτήρια που προσφέρονται και ανάλογα με τις φορές που συναντάται στο κείμενο της συγκεκριμένης σελίδας (βλ. Εικόνα 6).

Εικόνα 6: Αναζήτηση γλωσσικών στοιχείων στον φυλλομετρητή

iii. *Αναζήτηση συνομιλιών*: ο χρήστης μπορεί να επιλέξει συνομιλίες χρησιμοποιώντας έναν συνδυασμό κριτηρίων αναζήτησης (φύλο, ηλικία παιδιού, τόπος διαμονής, μονόγλωσσο/ δίγλωσσο παιδί κ.ά.)

ΑΠΟΜΑΓΝΗΤΟΦΩΝΗΣΗ ΥΛΙΚΟΥ	ΦΥΛΟ ΝΗΠΙΟΥ 1	ΗΛΙΚΙΑ ΝΗΠΙΟΥ 1	ΦΥΛΟ ΝΗΠΙΟΥ 2	ΗΛΙΚΙΑ ΝΗΠΙΟΥ 2	ΧΩΡΟΣ	ΤΟΠΟΣ ΔΙΑΜΟΝΗΣ	ΗΜΕΡΟΜΗΝΙΑ	ΔΙΑΡΚΕΙΑ	ΚΩΔΙΚΟΣ 1	ΚΩΔΙΚΟΣ 2
Ερευνητρια Ευ. Σκ.	Κορίτσι	3 ετών και 6 μηνών			Σπίτι νηπίου, αυτοκίνητο	Βόλος	December 17, 2014	10' 56"	1f3.6MLU2014	
Ερευνητρια Μα. Μπ.	Αγόρι	5 ετών και 5 μηνών			Σπίτι νηπίου	Βόλος	January 11, 2015	5' 02"	2m5.5MLU2015	
Ερευνητρια Βα. Γε.	Αγόρι	4 ετών και 6 μηνών			Σπίτι νηπίου	Αγιά Λάρισα	April 13, 2015	10' 35"	3m4.6MLR2015	

Εικόνα 7: Επιλογή συνομιλιών με κριτήρια αναζήτησης

iv. *Χρήσιμοι σύνδεσμοι*: σύνδεσμοι για προφορικά και γραπτά σώματα κειμένων, παιδικού και ενήλικου λόγου, από την ελληνική αλλά και άλλες γλώσσες.

*Διαθεσιμότητα*: το ΠΣΚΕΠ προσφέρει ελεύθερη πρόσβαση στα δεδομένα του στη διεύθυνση <http://gcs1.ece.uth.gr>.

### 3 Εφαρμογές, Προοπτικές και Περιορισμοί

Ο αρχικός στόχος δημιουργίας του ΠΣΕΠ ήταν η διεκδίκηση της ποιοτικής έρευνας του παιδικού λόγου: φυσικά, ένα μέρος του Σώματος μπορεί να αξιοποιηθεί και ποσοτικά. Η δυνατότητα που δίνεται είναι να μελετηθεί ο λόγος των νηπίων είτε από την άποψη των συνομιλιακών ικανοτήτων, είτε από την άποψη της συνοχής και συνεκτικότητας, εκεί όπου ο λόγος είναι συνεχής ή/ και περιλαμβάνει μικρές αφηγήσεις και περιγραφές: είναι επίσης δυνατή η χρήση των δεδομένων για διαγλωσσικές συγκρίσεις. Πέρα από τη μελέτη των συνομιλιών, δίνεται η δυνατότητα να απομονώσει κανείς συγκεκριμένα γλωσσικά φαινόμενα, αλλά και να τα μελετήσει σε σχέση με ορισμένες μεταβλητές (όπως φύλο ή ηλικία). Ιδιαίτερη σημασία έχουν τα δεδομένα από τον αυθόρμητο λόγο για τη μελέτη φαινομένων όπως οι παιδικές αποκλίσεις και οι νεολογισμοί (Μότσιου, 2016). Ένα αυθεντικό γλωσσικό υλικό μπορεί να αξιοποιηθεί τόσο ερευνητικά, και μάλιστα για διάφορες ειδικότητες και σκοπούς (γλωσσολόγοι, ψυχολόγοι, λογοθεραπευτές κτλ) όσο και παιδαγωγικά, ως αυθεντικό γλωσσικό υλικό/ πηγή δεδομένων στη διδασκαλία.

Η διεύθυνση των δυνατοτήτων της αναζήτησης και η επισημείωση των δεδομένων είναι μια προέκταση του έργου που θα διευκόλυνε σημαντικά την αναζήτηση μεμονωμένων γλωσσικών στοιχείων: τόσο η εξειδικευμένη αναζήτηση όσο και ο συνεχής εμπλουτισμός του ΠΣΚΕΠ αποτελούν μελλοντικές προοπτικές, ωστόσο οι στόχοι αυτοί αναπροσαρμόζονται πάντα σε σχέση τη διαθεσιμότητα των ερευνητικών προγραμμάτων και πόρων.

#### 4 Βιβλιογραφικές Αναφορές

- Behrens, H. (2008). *Corpora in Language Acquisition Research: History, Methods, Perspectives*. Amsterdam/Philadelphia: John Benjamins.
- CHILDES (Child Language Exchange System). Διαθέσιμο στο: <https://chilides.talkbank.org/> [25/03/2020]
- Diessel, H. 2009. Corpus linguistics and first language acquisition. In A. Lüdeling & M. Kytö (eds) *Corpus Linguistics. An International Handbook*. Volume 2. Berlin/New York: Walter de Gruyter, 1197-1212.
- Goutsos, D. & Fragaki, G. (2015). *Introduction to Corpus Linguistics* [Γούτσος, Δ. & Φραγκάκη, Γ., *Εισαγωγή στη Γλωσσολογία Σωμάτων Κειμένων*]. Athens: SEAB. Διαθέσιμο στο: <http://repository.kallipos.gr/handle/11419/1932> [25/03/2020]
- MacWhinney, B. (2000). *The CHILDES project: tools for analysing talk*. 3<sup>rd</sup> ed., Mahwah, NJ: Lawrence Erlbaum Associates.
- Μότσιου, Ε. (2017). Νεολογικές κατασκευές στον αυθόρμητο λόγο ρωσόφωνων και ελληνόφωνων παιδιών: μια πρώτη προσέγγιση. *Γλωσσολογία* 25, 19-31. Διαθέσιμο στο: <http://glossologia.phil.uoa.gr/node/136> [25/03/2020]
- Pavlidou, Th.-S. (2012). The Corpus of Spoken Greek: goals, challenges, perspectives. *LREC Proceedings, Workshop 18 (Best Practices for Speech Corpora in Linguistic Research)*, 23-28.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Wynne, M. (ed.) (2005). *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books. Διαθέσιμο στο [http://icar.cnrs.fr/ecole\\_thematique/contaci/documents/Baude/wynne.pdf](http://icar.cnrs.fr/ecole_thematique/contaci/documents/Baude/wynne.pdf) [25/03/2020]

#### 5 Ευχαριστίες

Ιδιαίτερες ευχαριστίες οφείλονται στους πρωτοετείς φοιτητές και φοιτήτριες του Παιδαγωγικού Τμήματος Προσχολικής Εκπαίδευσης του Πανεπιστημίου Θεσσαλίας που συνέλεξαν σε πρώτη φάση το υλικό και συνεχίζουν να συμβάλλουν στη μελλοντική διεύρυνση του ΠΣΚΕΠ.

Το έργο υλοποιήθηκε με χρηματοδότηση της Επιτροπής Ερευνών (Ειδικός Λογαριασμός Κονδυλίων Έρευνας) του Πανεπιστημίου Θεσσαλίας κατά το ακαδημαϊκό έτος 2019-20.