



ParlaCLARIN Workshop: Creating and Using Parliamentary Corpora

Miyazaki, 7 May 2018

Exploring the Political Agenda of the Greek Parliament Plenary Sessions

Dimitris Gkoumas, **Maria Pontiki**, Konstantina Papanikolaou, and Haris Papageorgiou
ATHENA Research & Innovation Centre/Institute for Language & Speech Processing
(ATHENA RIC/ILSP), Greece.



“Computational Science and Technologies: Data, Content and Interaction” (MIS 5002437)

“APOLLONIS, Greek Infrastructure for Digital Arts and Humanities, Language Research and Innovation” (MIS 5002738)

www.athenarc.gr

ParlaCLARIN@LREC2018, 7 May 2018, Miyazaki, Japan

Parliament plenary sessions: Valuable Data Sources

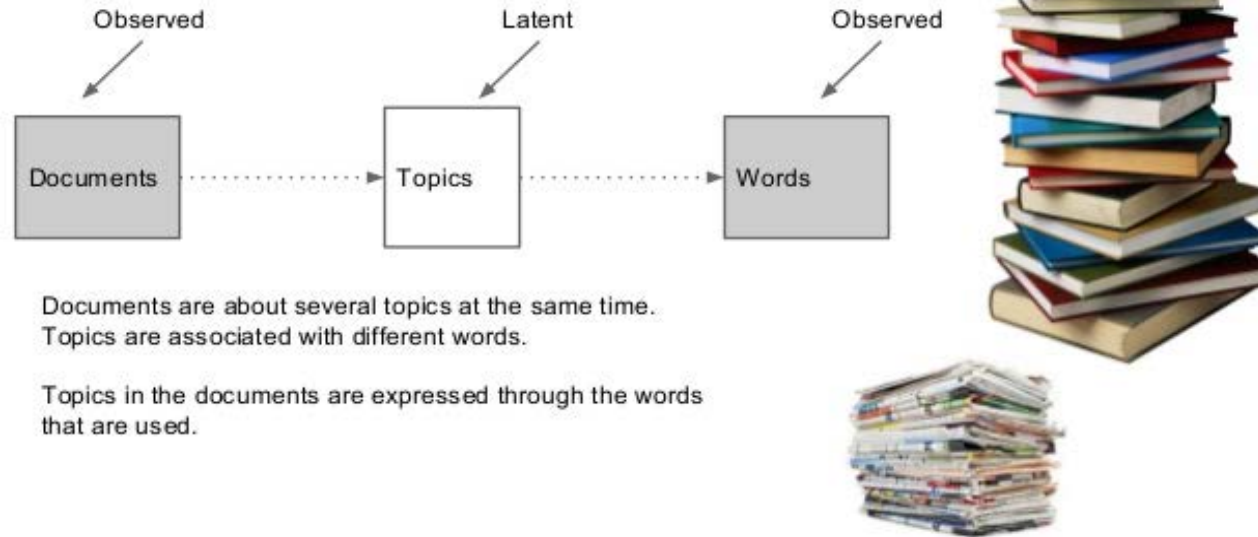


- Capturing the way political and policy agendas are shaped and evolve over time.
- Reflecting the stance and the political behavior under which individual Parliament Members and political groups react and act over time.



Parliament plenary sessions & Topic Modelling

Goal of Topic Modeling



Insights about:

- the way in which the European Parliament reacts to exogenous events like the Eurocrisis (Greene and Gross, 2017)
- the political attention of individual U.S. Senators (Quinn et al., 2010)



Image source: <https://theintelligenceofinformation.wordpress.com/2016/12/06/topic-modeling-latent-dirichlet-allocation-vs-correlation-explanation-alternative/>

www.athenarc.gr

Exploring the Political Agenda of the Greek Parliament Plenary Sessions

Methodology: Overview



Data Collection



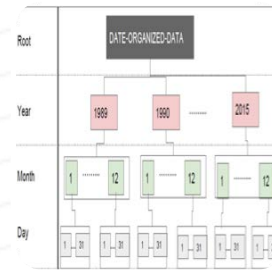
Web Scraping Technology



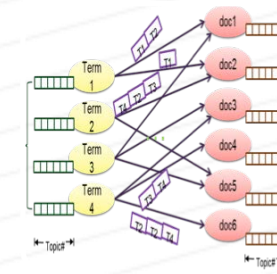
Structured Data



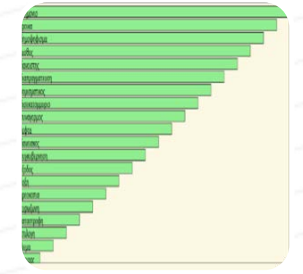
Preprocessing



Organization



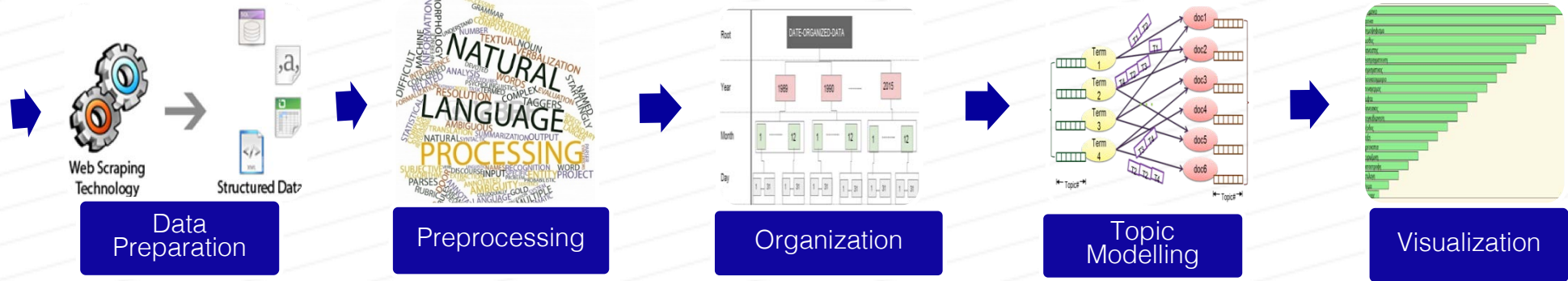
Topic Modelling



Visualization



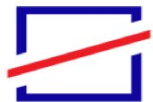
Data Collection



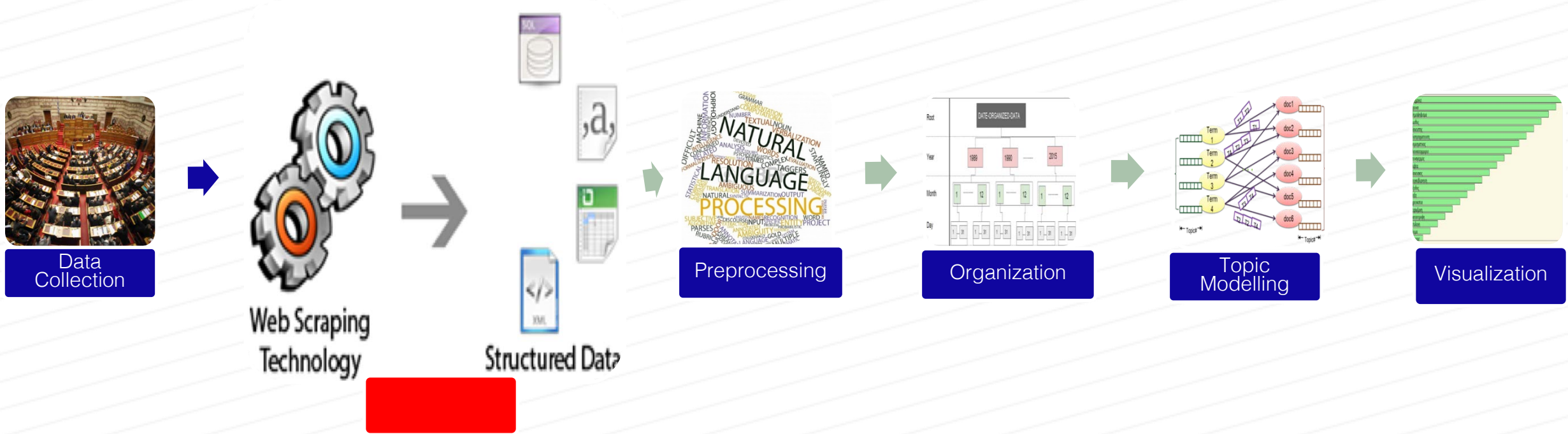
1989 to 2015 proceedings

→ 4356 plenary sessions

→ 1.063.546 unique speeches

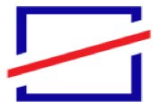


Data Preparation



Different formats (txt/doc/docx/pdf/jpeg)

→ structured and machine-readable xml version



Preprocessing



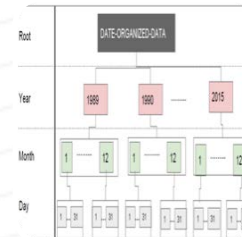
Data Collection



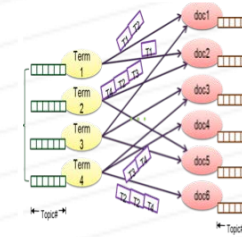
Web Scraping Technology



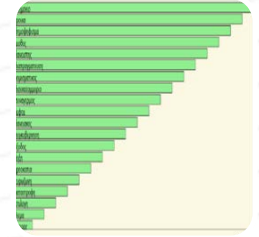
Structured Data



Organization

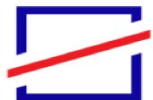


Topic Modelling



Visualization

- ✓ Segment annotation: the start/end point of each speech
- ✓ Temporal annotation: the date of each speech
- ✓ Speaker annotation: the name of each speaker and his/her political affiliation



- ✓ POS tagging

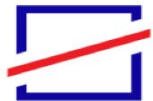
Data Organization



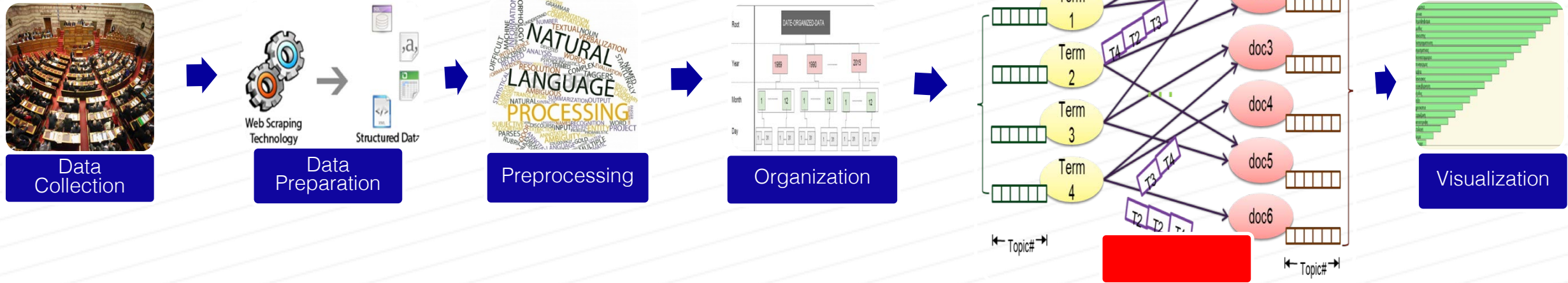
✓ time windows of fixed duration → 1 month

✓ 283 windows for the time period from 1989 to 2015

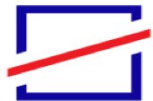
→ discover topics within the monthly proceedings (e.g. segments) & monitor their dynamics over time



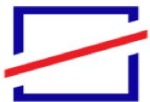
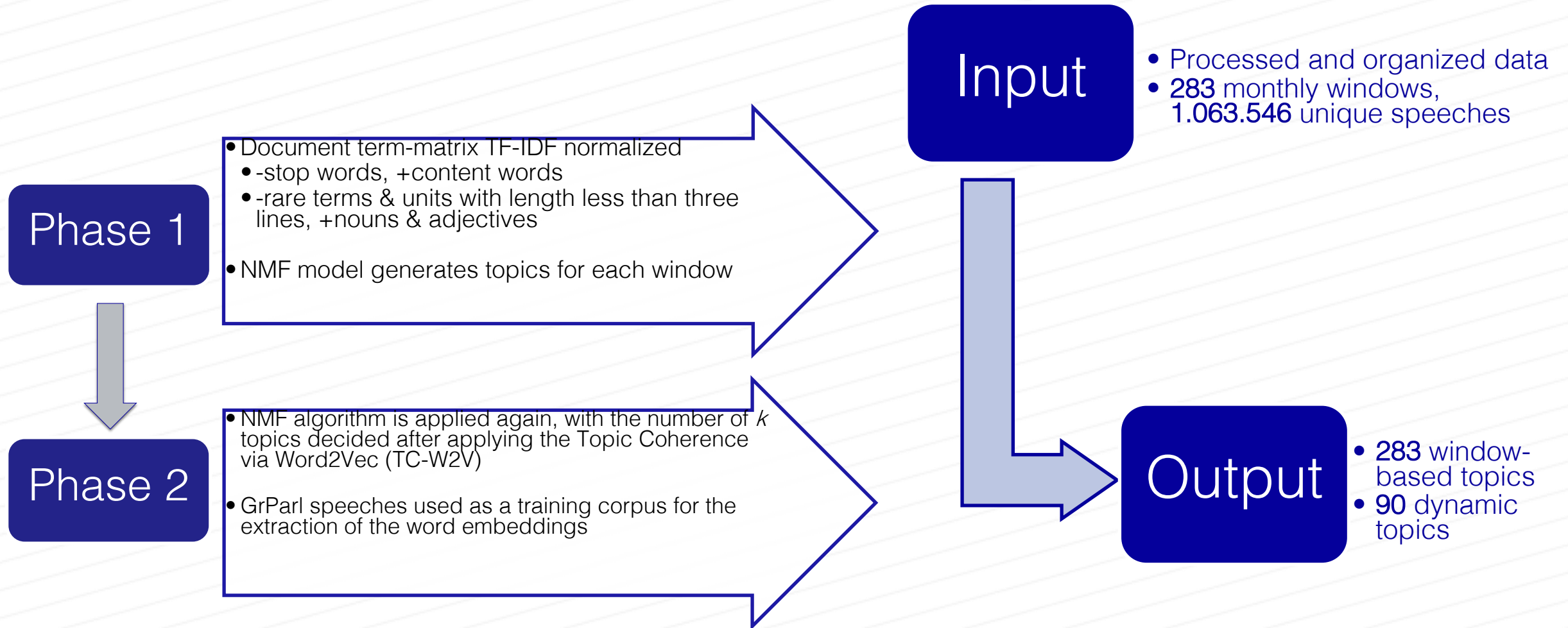
Topic Modelling



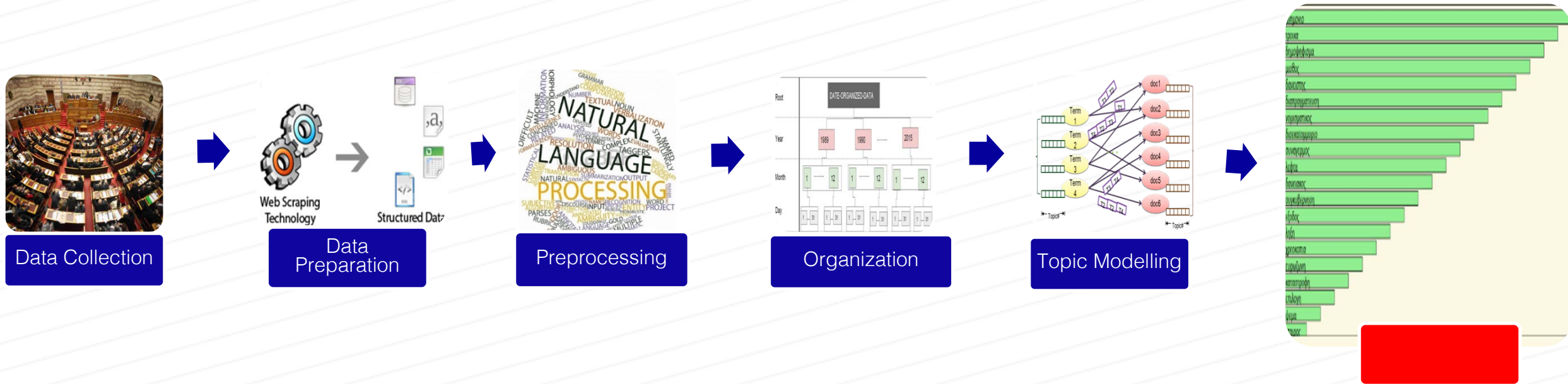
- ❑ Dynamic Non-negative Matrix Factorization (NMF) approach (Greene and Cross, 2017).
- ❑ After identifying a certain number of topics in each 1-month-window, the ones which are semantically similar are grouped together as “dynamic topics”, rather than identifying topics from the beginning.



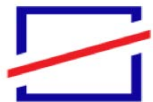
Dynamic NMF



Visualization



- ✓ By topic
- ✓ By time
- ✓ Search function



Greek Parliament Plenary Sessions presented by topic

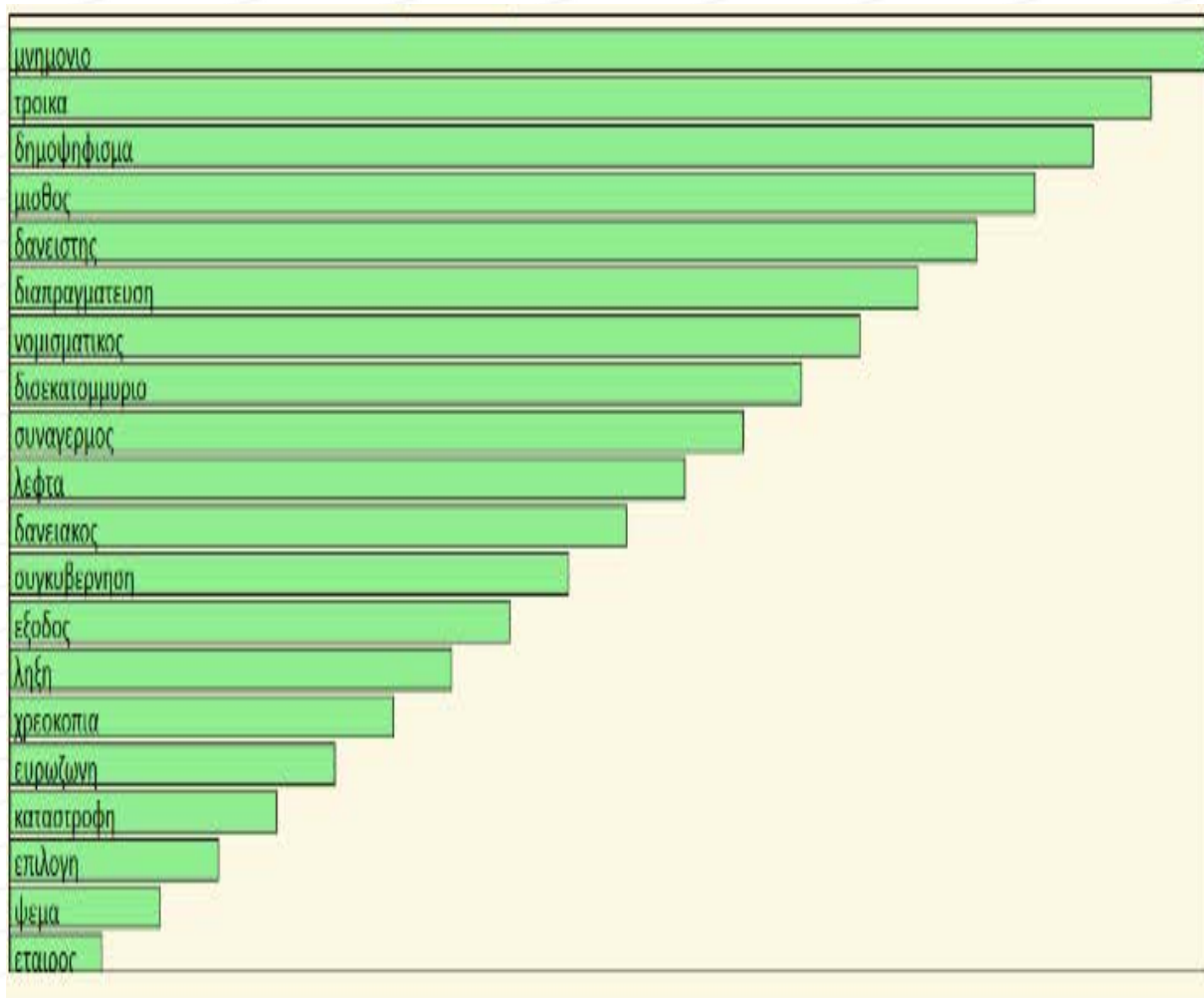
Show 50 entries

Search:

ID	Terms	#Parties	#Regions	Frequency
D01	κεφαλαιο, περιουσια, οδηγια, αμοιβαιος, εργατικος, ταξη, ομιλος, νομοθεσια, κατευθυνση, αξιοποιηση, στρωμα, οριο, μετοχη, ιδιωτικοποιηση, ορος, κεφαλαιαγορα, επικουρικος, πλουτος, ακινητα, επιπεδο,	18	58	68
D02	τροπολογια, προσθηκη, περιεχομενο, λογιστηριο, τροποποιηση, αιτημα, παραταση, σειρα, εκπροθεσμος, τυπικος, αποδοχη, διορθωση, προθεσμια, αδικια, βραδυ, κυβερνητικος, δεσμευση, μισος, υπογραφη, ονομαστικος,	19	58	71
D03	δημοκρατια, τοπος, δημοκρατικος, εξουσια, στελεχος, επιχειρημα, διαμαρτυρια, σταση, κυβερνητικος, καιρος, εντυπωση, εποχη, διακυβερνηση, υπευθυνος, συνασπισμος, εκφραση, πατριδα, μελλον, αντυπαραθεση, ευκαιρια,	19	58	71
D04	υγεια, μοναδα, προνοια, φροντιδα, ιατρικος, παροχη, νοσηλευτικος, προληψη, θεραπεια, ελλειψη, προμηθεια, αλληλεγγυη, ιδιωτικος, επιπεδο, πληθυσμος, προτεραιοτητα, εντατικος, εξοπλισμος, περιστατικο, ηγεσια,	18	58	71
D05	αγροτης, εισοδημα, επιδοτηση, παραγωγη, υπαιθρος, ενισχυση, καλλιεργεια, κοστος, επαγγελμα, πληθυσμος, βαμβακι, κινητοποιηση, στρεμμα, μητρωο, χωριο, δραχμη, επιστροφη, καταβολη, πληρωμη, χωραφι,	17	58	71



20 most prominent words for the dynamic topic D79



Involved Parliament Members

Show 10 entries

Download Segments organized by party

Search:

rank	Speaker	Party	Region	Date
1	ΑΝΑΣΤΑΣΙΟΣ ΠΕΤΡΟΠΟΥΛΟΣ			16_10_2015
1	Φίλης Νικόλαος Αριστοτέλη	ΣΥΝΑΣΠΙΣΜΟΣ ΡΙΖΟΣΠΑΣΤΙΚΗΣ ΑΡΙΣΤΕΡΑΣ	Α' ΑΘΗΝΩΝ	06_04_2015
1	Γκόκας Χρήστος Ναπολέοντα	ΠΑ.ΣΟ.Κ.	ΑΡΤΗΣ	08_05_2014
1	Χαραλαμπίδου Δέσποινα Δημητρίου		ΘΕΣΣΑΛΟΝΙΚΗΣ	16_07_2013
1	Λαφαζάνης Παναγιώτης Γεωργίου		ΠΕΙΡΑΙΩΣ	10_02_2015
1	Μηταράκης Παναγιώτης (Νότης)		ΑΘΗΝΩΝ	17_01_2014
1	Τσίπρας Αλέξιος Παύλου		ΑΘΗΝΩΝ	07_12_2014
1	Τσίπρας Αλέξιος Παύλου		ΑΘΗΝΩΝ	10_10_2014
1	Καμμένος Παναγιώτης (Πάνος)		ΑΘΗΝΩΝ	18_12_2012
1	Κεφαλογιάννη Όλγα Ιωάννη	ΝΕΑ ΔΗΜΟΚΡΑΤΙΑ	Α' ΑΘΗΝΩΝ	06_08_2014

Κεφαλογιάννη Όλγα Ιωάννη, ΝΕΑ ΔΗΜΟΚΡΑΤΙΑ - Α' ΑΘΗΝΩΝ, 06_08_2014

Ευχαριστώ, κύριε Πρόεδρε.

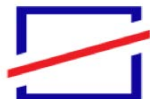
Κυρίες και κύριοι συνάδελφοι, έχουμε σήμερα μπροστά μας ένα σημαντικότατο κείμενο που αποκρυσταλλώνει τη θέληση της Κυβερνήσεως να συνεχίσει το έντονο μεταρρυθμιστικό της έργο.

Το παρόν σχέδιο νόμου ελάχιστα έχει να κάνει με τις επιταγές της τρόικα και περισσότερο με τις επιταγές της ελληνικής πραγματικότητας, μίας πραγματικότητας που προέκυψε έπειτα από δεκαετίες ανύπαρκτης αναπτυξιακής πολιτικής, μίας πραγματικότητας που προέκυψε έπειτα από δεκαετίες ανεύθυνης διασπάθισης του δημόσιου χρήματος, σαν να μην ήταν δικό μας, έπειτα από δεκαετίες κρατικοδίαιτης οικονομίας και υδροκεφαλισμού του κρατικού μηχανισμού.

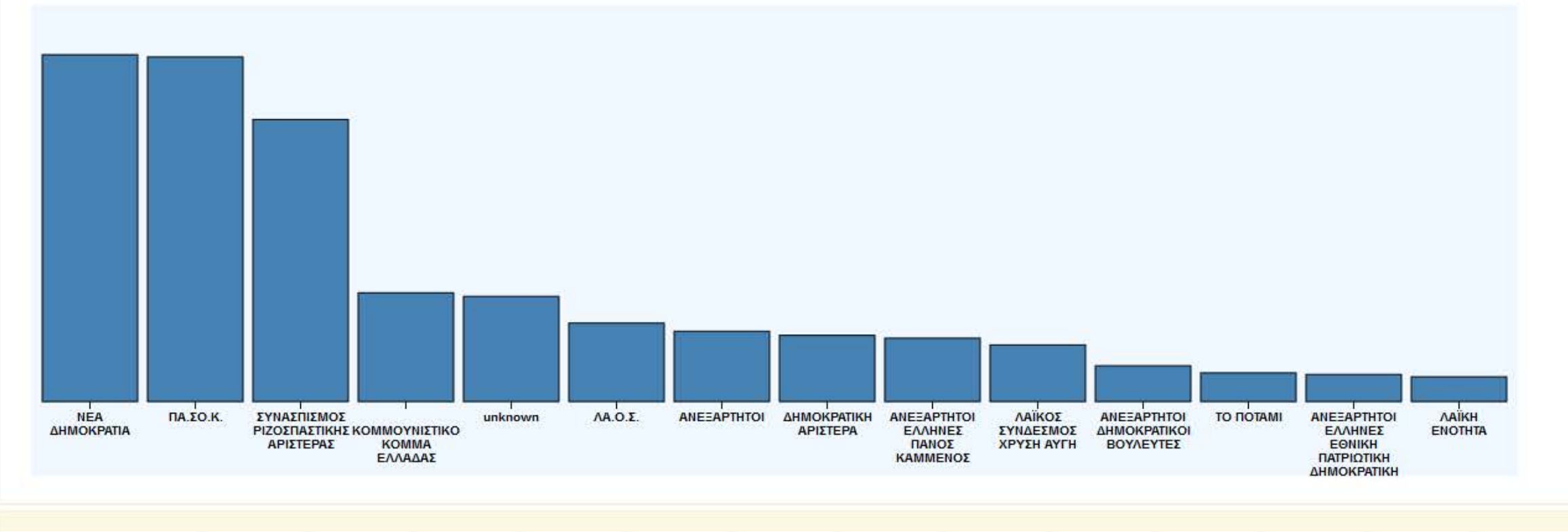
Στις δημαγωγίες της Αντιπολίτευσης η Κυβέρνηση απαντά με ένα προεκλεκτό κείμενο για την παραγωγική αναδινομιμία του ελληνικού κράτους.

Showing 1 to 10 of 800 entries

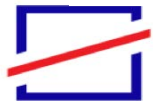
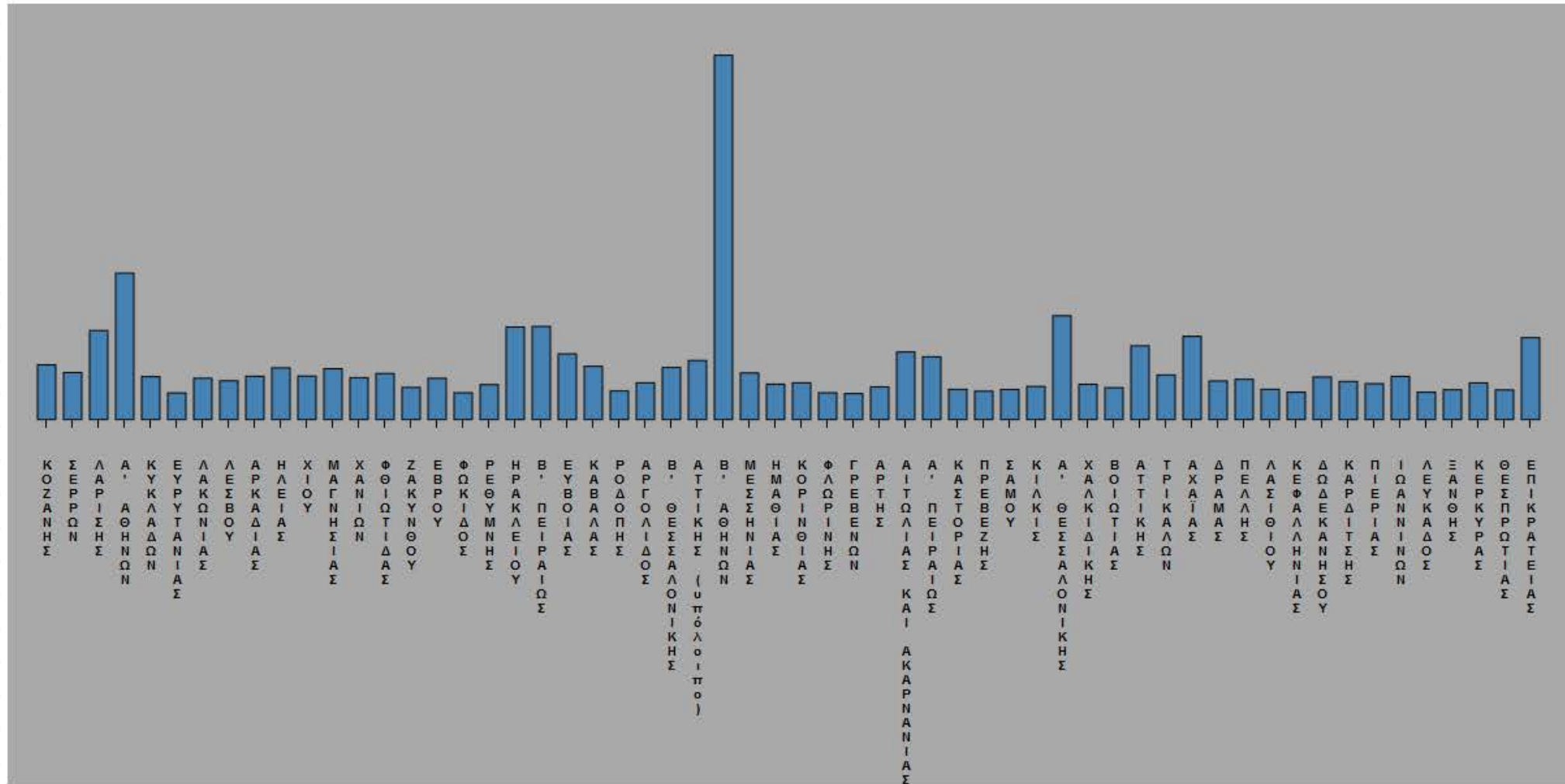
Previous 1 2 3 4 5 ... 80 Next



Contribution of each political party to a dynamic topic

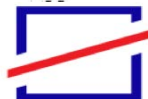


Contribution of each region to a dynamic topic



Most important window-topics related to a dynamic topic

Time Window	Topic ID	Terms	#Parties	#Politicians	#Regions	#Segments
2009_09	02	μνημονιο, επιχειρηση, εθνικος, διοικητης, αρμοδιοτητα, απασχοληση, προστασια, συμφωνια, συνεργασια, εδαφος, αδεια, νατοικος, συμμετοχη, προυποθεση, ορος, εξουσια, κυρωση, τριακοσιοι, περιεχομενο, συναγερμος,	5	37	20	140
2010_05	14	μετρο, συνταξη, νομισματικος, τιμη, χρεος, μισθος, μνημονιο, στηριξη, οικονομια, τραπεζα, αγωνας, συμφωνια, αγορα, μηχανισμος, δανειο, λυση, ελλειμμα, πατριδα, συμβαση, συνταξιουχος,	8	273	56	1523
2010_07	11	παιδι, μετρο, κοινωνια, οικογενεια, γενια, πολυτεκνος, γυναικα, εκατομμυριο, δουλεια, παιδεια, εθνικος, πληθυσμος, μνημονιο, οικονομια, μελλον, εργατικος, αγωνας, αγορα, λυση, πανεπιστημιο,	8	229	50	1465
2010_08	07	μνημονιο, μετρο, κοινωνια, κεφαλαιο, παραγωγη, αγωνας, εφοπλιστης, ναυτεργατης, τροικα, στρωμα, ανατροπη, τοπος, υπευθυνος, ταξη, εξουσια, εργατικος, κοστος, χρεοκοπια, χρεος, απελευθερωση,	8	65	31	281
2010_09	10	δισεκατομμυριο, χρεος, ελλειμμα, μνημονιο, πετρελαιο, επενδυση, λεφτα, μισθος, αναπτυξιακος, προυπολογισμος, τροικα, θερμανση, συνταξη, εθνικος, φορος, τιμη, οικονομια, μηχανισμος, ενισχυση, μετρο,	8	155	45	785
2010_10	14	μνημονιο, μετρο, ανεργια, επιλογη, προοπτικη, ανεργος, συμβαση, συνεργασια, φτωχεια, υπηρεσια, οικονομια, στηριξη, ορος, ελεγχος, εναλλακτικος, συνταξιουχος, βαρος, εξηνητα, μικρος, τροικα,	8	164	44	800
2010_12	01	προυπολογισμος, ελλειμμα, χρεος, δισεκατομμυριο, δημοκρατια, μετρο, μνημονιο, εσοδο, οικονομια, εκατομμυριο, φορος, μισθος, φορολογικος, θυσια, συνταξη, σπαταλη, επενδυση, αναπτυξιακος, προσαρμογη, νομισματικος,	8	285	57	1187



Greek Parliament Plenary Sessions presented by time

Show 50 entries

➔ Search:

Time Window	#Topics	#Terms	#Parties	#Politicians	#Segments
1989_07	11	174	5	95	642
1989_09	25	360	5	177	1633
1989_10	10	154	5	82	604
1989_11	23	34	5	23	53
1989_12	10	146	6	76	309
1990_01	10	164	5	133	567
1990_02	12	193	4	119	768
1990_03	21	223	4	119	455
1990_04	15	141	5	37	149
1990_05	19	288	5	108	612



Discussion

Contribution

Transformation of the Greek Parliament plenary sessions minutes into a structured and machine readable corpus

✓ *For the first time available for Natural Language Processing and Text Mining approaches*

Insightful navigation across the Greek Parliament corpus

✓ *Topic based presentation*

✓ *Time-window based presentation*

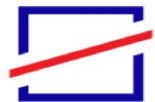
✓ *Search function*

Future Work

Qualitative analysis of the extracted dynamic topics & assignment of descriptive labels

Exploration of the evolution of selected topics (e.g. refugee crisis, Eurocrisis) & comparison with the European Parliament and/or other European countries

Enrichment with more insights e.g. Quotations, Sentiment



EΥΧΑΡΙΣΤΩ/**GRÀCIES**/HVALA/DĚKUJI/TAK/**DANK**
JEWEL/AITÄH/KIITOS/MERCI/**DANKE**/KÖSZNÖNÖ/
GRAZIE/PALDIES/**AČIŪ**/GRAZZI/TAKK/DZIEKUJE
OBRIGADO/MULTUMESC/**СПАСИБО**/ХВАЛА
HVALA/**БЛАГОДАРЯ**/THANKYOU/TAK/GRACIAS
/KIITOS/TACK/TEŞEKKÜREDERİM/**СПАСИБИ**/
JUFALMINDERIT/**ΕΥΧΑΡΙΣΤΩ**/DANKJEWEL/TAK
TACK/**GRAZZI**/DANKJEWEL/MULTUMESC/AITÄH
KÖSZNÖNÖ/СПАСИБО/ХВАЛА/AČIŪ/THANKYOU



Exploring the Political Agenda of the Greek Parliament Plenary Sessions

Dimitris Gkoumas, Maria Pontiki, Konstantina Papanikolaou, and Haris Papageorgiou

ATHENA-Research & Innovation Centre in Information, Communication and Knowledge Technologies (ATHENA RIC), Greece.



“Computational Science and Technologies: Data, Content and Interaction” (MIS 5002437)

“APOLLONIS, Greek Infrastructure for Digital Arts and Humanities, Language Research and Innovation” (MIS 5002738)

www.athenarc.gr

ParlaCLARIN@LREC2018, 7 May 2018, Miyazaki, Japan