

Introduction to Computational Linguistics, Natural Language Processing, and Large Language Models. Is it possible to work together?

Presentation @ Faculty of Foreign Languages of UT
June 5, 2025, Tirana
Updated October 15, 2025

Nikitas N. Karanikolas

Dept. of Informatics and Computer Engineering

University of West Attica

<https://users.uniwa.gr/nnk/>

nnk@uniwa.gr

Short Definitions to CL, NLP, LLM – 1/3

- **Linguistics** is the scientific study of language. It spans both theoretical and practical applications. **Theoretical linguistics** is concerned with understanding the universal and fundamental nature of language and developing a general theoretical framework for describing it. **Applied linguistics** seeks to utilize the scientific findings of the study of language for practical purposes, such as developing methods of improving language education and literacy. Linguistics emerged from the field of philology. Linguistics is also related to the philosophy of language, stylistics, rhetoric, semiotics, lexicography, and translation.
- **Computational Linguistics (CL)** is an interdisciplinary field concerned with the computational modelling of natural language, as well as the study of appropriate computational approaches to linguistic questions. Computational linguistics draws upon linguistics, computer science, artificial intelligence, logic, philosophy, cognitive science.

Short Definitions to CL, NLP, LLM – 2/3

- **Natural Language Understanding (NLU)** is the effort, methods, tools and techniques to transform the human/natural language text to some structured form (like Predicate Calculus, FRL, KL-ONE, Case Grammars, etc) and handle it by computers.
- **Natural Language Generation (NLG)** is the reverse. It is a software process that produces natural written or spoken text from structured and unstructured data.
- **Natural Language Processing (NLP)** is a broader term. It is a branch of Artificial Intelligence (AI) which enables computers to analyze and synthesize natural (human) language and speech. It includes both NLU and NLG and more others:
 - Summarization,
 - Question Answering,
 - Machine Translation,
 - Text Classifications,
 - Sentiment Analysis,
 - Named Entity Recognition,

Short Definitions to CL, NLP, LLM – 3/3

- **Large Language Models (LLMs)** is a software process that uses deep learning techniques and massively large data sets to learn how to understand and generate natural text
- The process of evolution of language models has gone under four main stages:
 - The first stage was in 1990s, where statistical models on n-gram languages were used both in NLP and Information Retrieval tasks
 - The next stage involved the use of neural networks, such as RNNs, to be replaced later by the more efficient bi-LSTM models, leading to the class of Pre-trained Language Models (PLMs)
 - However, those models' performance excelled when the model's size or the data size was significantly increased
 - This was the opening of today's Large Language Models (LLMs)

Detailed presentation of CL

- The CL field overlapped with artificial intelligence since the efforts in the United States in the 1950s to use computers to automatically translate texts from foreign languages, particularly Russian scientific journals, into English. Since rule-based approaches were able to make arithmetic (systematic) calculations much faster and more accurately than humans, it was expected that lexicon, morphology, syntax and semantics can be learned using explicit rules, as well. After the failure of rule-based approaches, the term Computational Linguistics is coined in order to distinguish the field from AI.
- CL is a very active disciplinary supported by the Association of Computational Linguistics (ACL) and its more well known events are the COLING Conferences

Detailed Presentation of NLG 1/2

- Systems, models and applications that can produce discourse (written text or oral speech) with meaning not recognized as artificial
- Stages: Planning, Microplanning, Realizing, Presenting (optional)
- Planning - decide what is:
 - the most interesting parts of input data (structured data),
 - the order of communicated ideas/facts,
 - the hidden rhetorical relations [cause, sequence, etc]
- Microplanning:
 - aggregation,
 - anaphora generation,
 - selection of lexical items,
 - decide the syntactic

Detailed Presentation of NLG 2/2

- Realizing morphologically and orthographically correct discourse:
 - inflection,
 - orthography,
 - ordering of adjectives.
- Presenting (optional stage):
 - Written text: Titles, Emphasis (e.g. bold), punctuation marks
 - Oral speeches: Intonation, Sentence Type (Affirmative, Negative, Imperative, Exclamatory, Interrogative)
- Prerequisite for NLG is NLU

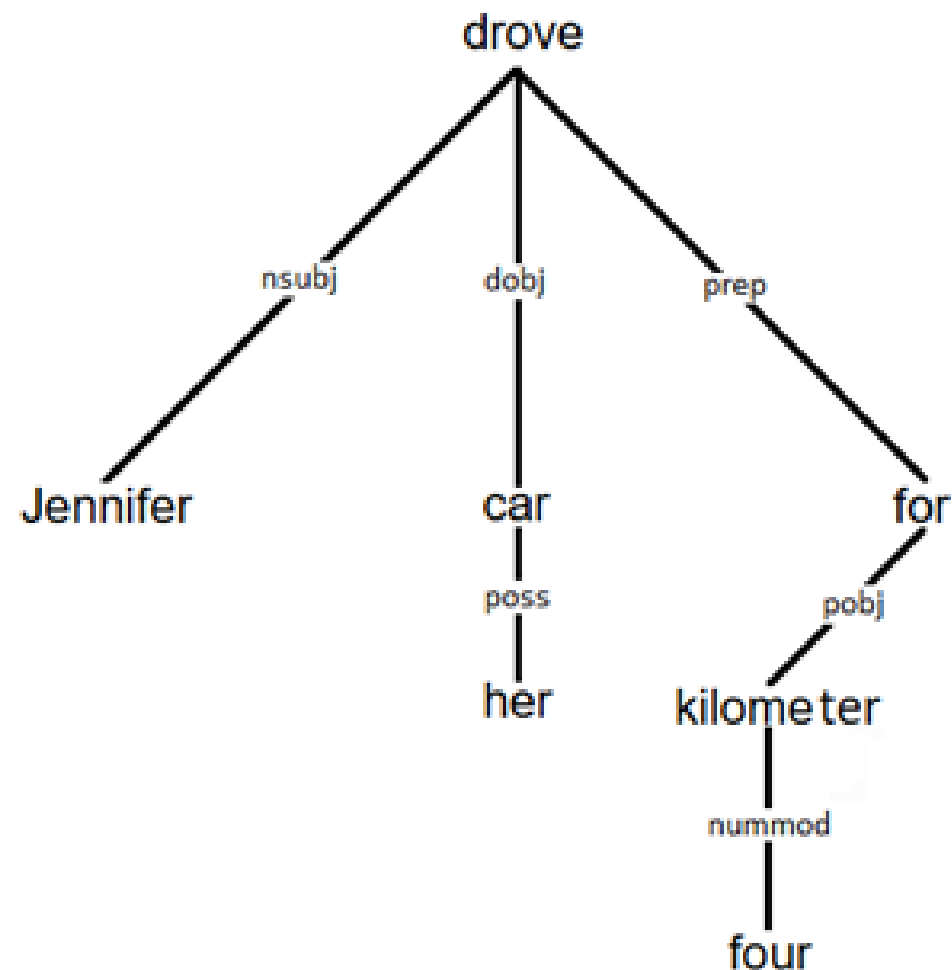
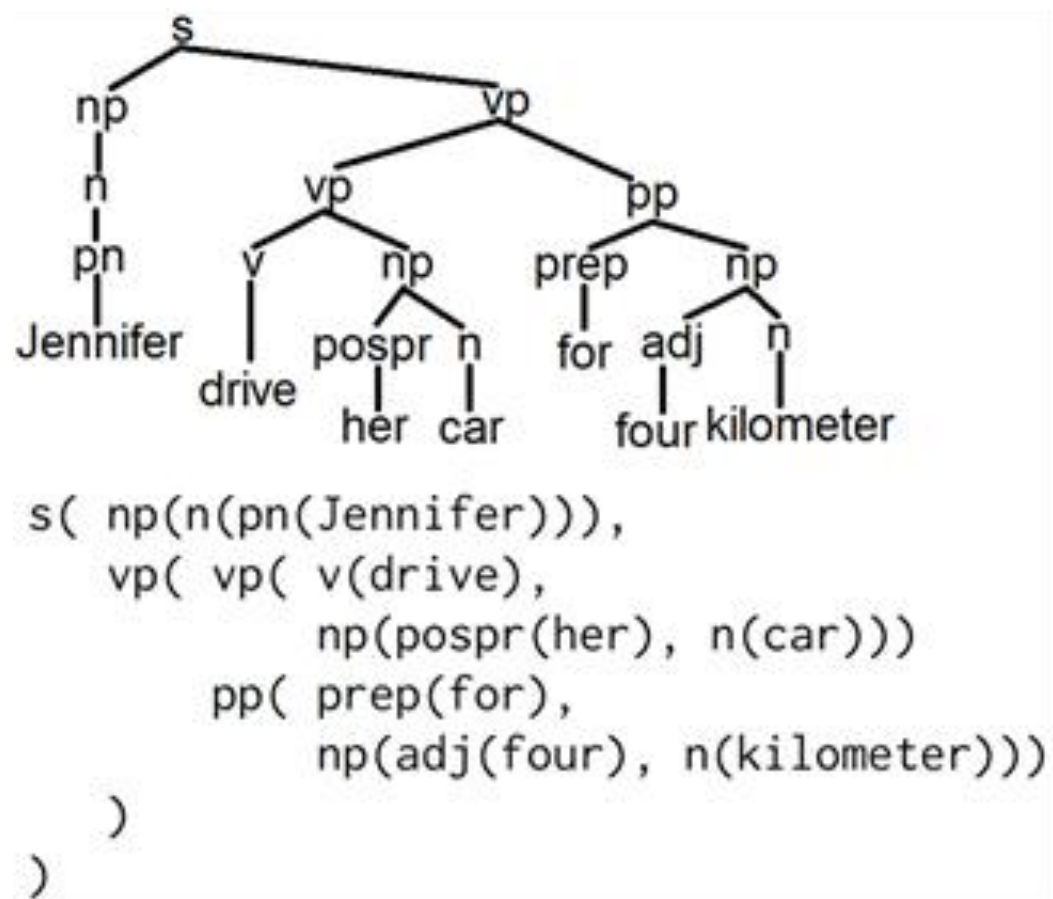
Detailed Presentation of NLU 1/7

- NLU is the translation of human/natural language text to some structured form (Predicate Calculus, FRL, KL-ONE, Case Grammars, etc)
- Natural Language Understanding follows (usually) the following tasks:
 - Tokenization,
 - Part of Speech Tagging,
 - Syntactic Analysis,
 - Structural Disambiguation (Resolution of Syntactic Ambiguity),
 - Word Sense Disambiguation,
 - Semantic Representation,
 - Anaphora Resolution,
 - Optional Tasks: Affective Computing, Discourse Analysis, Pragmatics.

NLU 2/7

Constituent Parsing versus Dependency Parsing

"Jennifer drove her car, for four kilometers"



NLU 3/7 Structural Disambiguation

(Resolution of Syntactic Ambiguity)

syntactic/parse (phrase structure) trees (in nested lists form)
resulting from the sentence "Look at the dog with one eye."

```
s( np(n(you))
  vp( v(look),
      pp (p(at),np(det(the),n(dog))),
      pp (p(with),np(adj(one),n(eye)))
    )
  )
```

```
s( np(n(you))
  vp( v(look),
      pp ( p(at),
            np( det(the),
                  n(dog),
                  pp(p(with), np(adj(one),n(eye)))
                )
          )
    )
  )
```

Semantic representation

e.g. “Jennifer drove her car, for four kilometers”

DRIVE

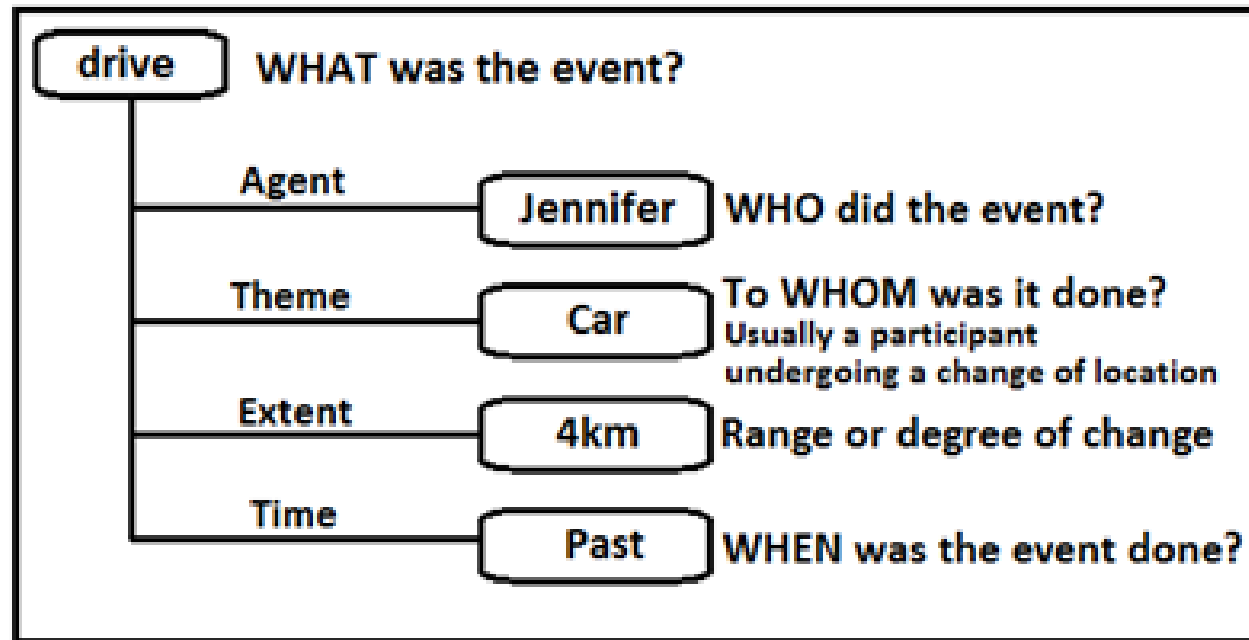
Acceptable Thematic Roles:

Agent, Theme, Source, Destination, Location, Extent, Duration, Time

Selectional Restrictions:

.....

Jennifer drove her car for four kilometers



NLU 5/7

Thematic Roles according to verbnet

Table 2: Thematic roles and example classes that use them

Actor:	used for some communication classes (e.g., Chitchat-37.6, Marry-36.2, Meet-36.2) when both arguments can be considered symmetrical (pseudo-agents).
Agent:	generally a human or an animate subject. Used mostly as a volitional agent, but also used in VerbNet for internally controlled subjects such as forces and machines.
Asset:	used for the Sum of Money Alternation, present in classes such as Build-26.1, Get-13.5.1, and Obtain-13.5.2 with 'currency' as a selectional restriction.
Attribute:	attribute of Patient/Theme refers to a quality of something that is being changed, as in (The price)att of oil soared. At the moment, we have only one class using this role Calibratable cos-45.6 to capture the Possessor Subject Possessor-Attribute Factoring Alternation. The selectional restriction 'scalar' (defined as a quantity, such as mass, length, time, or temperature, which is completely specified by a number on an appropriate scale) ensures the nature of Attribute.
Beneficiary:	the entity that benefits from some action. Used by such classes as Build-26.1, Get-13.5.1, Performance-26.7, Preparing-26.3, and Steal-10.5. Generally introduced by the preposition 'for', or double object variant in the benefactive alternation.
Cause:	used mostly by classes involving Psychological Verbs and Verbs Involving the Body.
Location, Destination, Source:	used for spatial locations.
Destination:	end point of the motion, or direction towards which the motion is directed. Used with a 'to' prepositional phrase by classes of change of location, such as Banish-10.2, and Verbs of Sending and Carrying. Also used as location direct objects in classes where the concept of destination is implicit (and location could not be Source), such as Butter-9.9, or Image impression-25.1.
Source:	start point of the motion. Usually introduced by a source prepositional phrase (mostly headed by 'from' or 'out of'). It is also used as a direct object in such classes as Clear-10.3, Leave-51.2, and Wipe instr-10.4.2.
Location:	underspecified destination, source, or place, in general introduced by a locative or path prepositional phrase.
Experiencer:	used for a participant that is aware or experiencing something. In VerbNet it is used by classes involving Psychological Verbs, Verbs of Perception, Touch, and Verbs Involving the Body.
Extent:	used only in the Calibratable-45.6 class, to specify the range or degree of change, as in The price of oil soared (10%)ext. This role may be added to other classes.
Instrument:	used for objects (or forces) that come in contact with an object and cause some change in them. Generally introduced by a 'with' prepositional phrase. Also used as a subject in the Instrument Subject Alternation and as a direct object in the Poke-19 class for the Through/With Alternation and in the Hit-18.1 class for the With/Against Alternation.
Material and Product:	used in the Build and Grow classes to capture the key semantic components of the arguments. Used by classes from Verbs of Creation and Transformation that allow for the Material/Product Alternation.
Material:	start point of transformation.
Product:	end result of transformation.
Patient:	used for participants that are undergoing a process or that have been affected in some way. Verbs that explicitly (or implicitly) express changes of state have Patient as their usual direct object. We also use Patient1 and Patient2 for some classes of Verbs of Combining and Attaching and Verbs of Separating and Disassembling, where there are two roles that undergo some change with no clear distinction between them.
Predicate:	used for classes with a predicative complement.
Recipient:	target of the transfer. Used by some classes of Verbs of Change of Possession, Verbs of Communication, and Verbs Involving the Body. The selection restrictions on this role always allow for animate and sometimes for organization recipients.
Stimulus:	used by Verbs of Perception for events or objects that elicit some response from an experiencer. This role usually imposes no restrictions.
Theme:	used for participants in a location or undergoing a change of location. Also, Theme1 and Theme2 are used for a few classes where there seems to be no distinction between the arguments, such as Differ-23.4 and Exchange-13.6 classes.
Time:	class-specific role, used in Begin-55.1 class to express time.
Topic:	topic of communication verbs to handle theme/topic of the conversation or transfer of message. In some cases, like the verbs in the Say-37.7 class, it would seem better to have 'Message' instead of 'Topic', but we decided not to proliferate the number of roles.

<https://verbs.colorado.edu/verbnet/> --> καρτέλα (tab) "About VerbNet" --> scroll down

Detailed Presentation of NLU 6/7

Πρόταση	Verb	Agent	Destination	Co-Agent	Instrument	Patient	Σχόλιο
Ο Παναγιώτης πήγε στο θέατρο με το λεωφορείο	πηγαίνω	Παναγιώτης	θέατρο		λεωφορείο		
	Indicative, Aorist, Singular, 3rd	Male, Singular, Nominative	Neutral, Singular, Nominative		Neutral, Singular, Nominative		
		Ο Παναγιώτης NP(DET(Ο), PROPN(Παναγιώτης))	στο θέατρο PP(PREP(στο), NOUN(θέατρο))		με το λεωφορείο PP(PREP(με), DET(το), NOUN(λεωφορείο))		αντι για ADP χρησιμοποιώ PREP (πρόθεση, preposition)
Ο Παναγιώτης πήγε στο θέατρο με τη Δήμητρα	πηγαίνω	Παναγιώτης	θέατρο	Δήμητρα			
	Indicative, Aorist, Singular, 3rd	Male, Singular, Nominative	Neutral, Singular, Nominative	Female, Singular, Nominative			
		Ο Παναγιώτης NP(DET(Ο), PROPN(Παναγιώτης))	στο θέατρο PP(PREP(στο), NOUN(θέατρο))	με τη Δήμητρα PP(PREP(με), DET(τη), PROPN(Δήμητρα))			
Ο Παναγιώτης πήγε τη Δήμητρα στο ξενοδοχείο (αυτό το παράδειγμα δεν υπάρχει στις από 1-6-2024 σημειώσεις μου)	πηγαίνω	Παναγιώτης	ξενοδοχείο			Δήμητρα	Ερώτηση: Γιατί η φράση 'τη Δήμητρα' τοποθετείται στο σημασιολογικό ρόλο Patient και όχι στον Co-Agent;
	Indicative, Aorist, Singular, 3rd	Male, Singular, Nominative	Neutral, Singular, Nominative			Female, Singular, Nominative	Selectional restriction Semantic role Co-Agent: Συνήθως PP με PREP 'με'.
		Ο Παναγιώτης	στο ξενοδοχείο			τη Δήμητρα	Selectional restriction Semantic role Patient: Συνήθως NP σε θέση Direct Object (Ονοματική Φράση - NP - που ακολουθεί το ρήμα).
		NP(DET(Ο), PROPN(Παναγιώτης))	PP(PREP(στο), NOUN(ξενοδοχείο))			NP(DET(τη), PROPN(Δήμητρα))	Απάντηση: Γιατί η φράση 'τη Δήμητρα' είναι NP και όχι PP και ως εκ τούτου δεν έχει την πρόθεση 'με' ή άλλη ανάλογη. Επιπλέον είναι Direct Object στο ρήμα.

Detailed Presentation of NLU 7/7

Πρόταση	Verb	Agent	Destination	Co-Agent	Instrument	Patient	Σχόλιο
Ο Παναγιώτης πήγε στο θέατρο με το προαστιακό λεωφορείο	πηγαίνω	Παναγιώτης	θέατρο		λεωφορείο		
	Indicative, Aorist, Singular, 3rd	Male, Singular, Nominative	Neutral, Singular, Nominative		Neutral, Singular, Nominative		
		Ο Παναγιώτης	στο θέατρο		με το προαστιακό λεωφορείο		
		NP(DET(Ο), PROPΝ(Παναγιώτης))	PP(PREP(στο), NOUN(θέατρο))		PP(PREP(με), DET(το), ADJ(προαστιακό), NOUN(λεωφορείο))		
Ο Παναγιώτης πήγε στο θέατρο με την όμορφη Δήμητρα	πηγαίνω	Παναγιώτης	θέατρο	Δήμητρα			
	Indicative, Aorist, Singular, 3rd	Male, Singular, Nominative	Neutral, Singular, Nominative	Female, Singular, Nominative			
		Ο Παναγιώτης	στο θέατρο	με την όμορφη Δήμητρα			
		NP(DET(Ο), PROPΝ(Παναγιώτης))	PP(PREP(στο), NOUN(θέατρο))	PP(PREP(με), DET(την), ADJ(όμορφη), PROPΝ(Δήμητρα))			
Ο Παναγιώτης πήγε στο θέατρο με την αγαπημένη του θεία	πηγαίνω	Παναγιώτης	θέατρο	θεία			
	Indicative, Aorist, Singular, 3rd	Male, Singular, Nominative	Neutral, Singular, Nominative	Female, Singular, Nominative			
		Ο Παναγιώτης	στο θέατρο	με την αγαπημένη του θεία			
		NP(DET(Ο), PROPΝ(Παναγιώτης))	PP(PREP(στο), NOUN(θέατρο))	PP(PREP(με), DET(την), ADJ(αγαπημένη), POSPRO(του), NOUN(θεία))			εδώ η λέξη 'του' δεν είναι DET αλλά είναι POSPRO (possessive pronoun)

Detailed Presentation of LLMs with usage examples 1/9

Μπορείς να μου δώσεις τους αρχικούς χρόνους του ρήματος πλέκω στην ενεργητική φωνή στα νέα ελληνικά, διακρίνοντας τον Μέλλοντα σε στιγμιαίο, εξακολουθητικό και συντελεσμένο;



Φυσικά! Στα Νέα Ελληνικά, οι αρχικοί χρόνοι του ρήματος **πλέκω** στην ενεργητική φωνή, με διάκριση του Μέλλοντα, παρουσιάζονται ως εξής:

- **Ενεστώτας:** πλέκω
- **Παρατατικός:** έπλεκα
- **Μέλλοντας:**
 - **Στιγμιαίος:** θα πλέξω
 - **Εξακολουθητικός:** θα πλέκω
 - **Συντελεσμένος:** θα έχω πλέξει
- **Αόριστος:** έπλεξα
- **Παρακείμενος:** έχω πλέξει
- **Υπερσυντέλικος:** είχα πλέξει



Αν χρειάζεσαι κάτι επιπλέον, πες μου! 😊

Detailed Presentation of LLMs with usage examples 2/9

Μπορείς να μου δώσεις τους αρχικούς χρόνους του ρήματος βλέπω στην ενεργητική φωνή στα νέα ελληνικά;



Βεβαίως! Οι αρχικοί χρόνοι του ρήματος **βλέπω** στην ενεργητική φωνή στα Νέα Ελληνικά είναι οι εξής:

- Ενεστώτας: βλέπω
- Παρατατικός: έβλεπα
- Μέλλοντας:
 - Εξακολουθητικός: θα βλέπω
 - Στιγμιαίος: θα δω
 - Συντελεσμένος: θα έχω δει
- Αόριστος: είδα
- Παρακείμενος: έχω δει
- Υπερσυντέλικος: είχα δει



Detailed Presentation of LLMs with usage examples 3/9

Μπορείς να μου βρεις τους αρχικούς χρόνους στα νέα ελληνικά του ρήματος τρώω;



Βεβαίως! Οι αρχικοί χρόνοι του ρήματος **τρώω** στην ενεργητική φωνή στα Νέα Ελληνικά είναι οι εξής:

- Ενεστώτας: τρώω
- Παρατατικός: έτρωγα
- Μέλλοντας:
 - Εξακολουθητικός: θα τρώω
 - Στιγμιαίος: θα φάω
 - Συντελεσμένος: θα έχω φάει
- Αόριστος: έφαγα
- Παρακείμενος: έχω φάει
- Υπερσυντέλικος: είχα φάει

Detailed Presentation of LLMs with usage examples 4/9

Failure
'τρωθώ' =
corrothion

Μπορείς να μου δώσεις τους αρχικούς χρόνους στα νέα ελληνικά του ρήματος τρώγομαι στην παθητική φωνή, διακρίνοντας το μέλλοντα σε εξακολουθητικό, στιγμιαίο και συντελεσμένο;



Βεβαίως! Το ρήμα "τρώγομαι" στην παθητική φωνή, με τη διάκριση του μέλλοντα σε εξακολουθητικό, στιγμιαίο και συντελεσμένο, έχει τους εξής αρχικούς χρόνους:

- Ενεστώτας: τρώγομαι
- Παρατατικός: τρωγόμεν
- Αόριστος: τραφώ
- Παρακείμενος: έχω τραφεί
- Εξακολουθητικός μέλλοντας: θα τρώγομαι
- Στιγμιαίος μέλλοντας: θα τρωθώ
- Συντελεσμένος μέλλοντας: θα έχω τραφεί
- Υπερσυντέλικος: είχα τραφεί



Detailed Presentation of LLMs with usage examples 5/9

Καλησπέρα, θα ήθελα να μετατρέψουμε κάποιες προτάσεις από την ενεργητική σύνταξη στην παθητική σύνταξη.

Καλησπέρα! Βεβαίως, γράψτε μου τις προτάσεις που θέλετε να μετατρέψουμε. 😊

Ο Ιορδάνης μετέφερε το γάλα στο ψυγείο.

Το γάλα μεταφέρθηκε στο ψυγείο από τον Ιορδάνη.

Detailed Presentation of LLMs with usage examples 6/9

Εχθές το βράδυ, ο Παναγιώτης έκανε τριανταπέντε τσουρέκια.

Εχθές το βράδυ, τριανταπέντε τσουρέκια έγιναν από τον Παναγιώτη.



Detailed Presentation of LLMs with usage examples 7/9

Μπορείς να μου μετατρέψεις την παρακάτω πρόταση από ενεργητική σύνταξη σε παθητική; Ο Παναγιώτης είχε δει τις σημειώσεις του καθηγητή όταν πήγε να τον συναντήσει.

Οι σημειώσεις του καθηγητή είχαν ειδωθεί από τον Παναγιώτη όταν πήγε να τον συναντήσει.

Υπάρχει κι άλλος τρόπος;

Ναι, μπορεί να διατυπωθεί και ως εξής:

Οι σημειώσεις του καθηγητή είχαν ήδη γίνει ορατές από τον Παναγιώτη όταν πήγε να τον συναντήσει.

Ή, αν θέλουμε να διατηρήσουμε τη φυσικότητα του λόγου:

Οι σημειώσεις του καθηγητή είχαν ήδη διαβαστεί από τον Παναγιώτη όταν πήγε να τον συναντήσει.

Detailed Presentation of LLMs with usage examples 8/9

Κλείδωσα την πόρτα πριν φύγω.

Η Μαρία έπλεξε ένα όμορφο κασκόλ για τη φίλη της.

Είδα ένα ενδιαφέρον ντοκιμαντέρ στην τηλεόραση.

Η Μαρία έφαγε το μεσημεριανό της γεύμα.

Ο δάσκαλος λέει μια ιστορία στους μαθητές του.

Ο πατέρας βγάζει την καρέκλα από την αποθήκη.

Η μητέρα βάζει τα ρούχα στο πλυντήριο.

Ο πεζός βρέχει τα πόδια του περπατώντας στη βροχή.

Ο στρατηγός στρέφει τις δυνάμεις του προς τη νέα κατεύθυνση.

Detailed Presentation of LLMs with usage examples 9/9

Ορίστε οι προτάσεις μετατρεπμένες σε παθητική σύνταξη:

1. Η πόρτα κλείδωσε πριν φύγω.
2. Ένα όμορφο κασκόλ πλέχτηκε από τη Μαρία για τη φίλη της.
3. Ένα ενδιαφέρον ντοκιμαντέρ είδε στην τηλεόραση.
4. Το μεσημεριανό γεύμα φαγώθηκε από τη Μαρία.
5. Μια ιστορία λέγεται από τον δάσκαλο στους μαθητές του.
6. Η καρέκλα βγαίνει από την αποθήκη από τον πατέρα.
7. Τα ρούχα μπαίνουν στο πλυντήριο από τη μητέρα.
8. Τα πόδια του πεζού βρέχονται περπατώντας στη βροχή.
9. Οι δυνάμεις του στρατηγού στρέφονται προς τη νέα κατεύθυνση.

Σωστό 1: Η πόρτα κλειδώθηκε [από εμένα] πριν φύγω.

Σωστό 3: Ένα ενδιαφέρον ντοκιμαντέρ ειδώθηκε [από εμένα] στην τηλεόραση.

Weaknesses of NLPs

- NLU/NLG systems **require skilled experts** to create rules
- There exist **numerous human languages** with varying grammar and syntax
- **Rules** are **numerous, complex, overlapping**, or **contradicting** and significant effort is needed for their maintenance
- **R&D** groups working on NLU/NLG worldwide are **isolated** and not sharing their techniques and accomplishments, due to interests / competition
- Since languages evolve, NLU/NLG systems **may become outdated**
- Handling **ambiguity** and **variability** of natural languages is **hard**
- NLU/NLG systems may suffer from **bias**, originating on opinions of their developers

Weaknesses of LLMs

- are prone to produce **misinformation**, **bias**, and **impolite** language resulting from their training data
- can generate fake, harmful, misleading, content, or propaganda, raising **ethical concerns**
- tend to ‘**hallucinate**’ and produce, with full conviction, high quality text which contains factually incorrect information
- since they rely on training data, **use potentially obsolete information** and may generate responses that are no longer accurate or relevant
- operate like **black boxes**, without the possibility to interpret / explain why they produced specific output
- are **not easily controllable** and when mistaken answers are identified, it is hard, if not impossible to diagnose and fix the error
- may handle sensitive and personal data and measures for **data protection** are needed
- are **costly to train**, since massive amounts of data, powerful computing systems and significant energy consumption are needed

Resolving Weaknesses

- Life attitude / Στάση ζωής
 - Are you planning to accept passively whatever changes that the exciting technology (and behind it the Companies) imposes. Young people prefer to be “?smart?” and use the technology to gain whatever they want, without recognizing the drawbacks and without asking why this is this way and not the other way.
 - There is no money any more. They are starting to exchange the plastic bank (credit, debit, prepaid) cards with other digital means (already done with smart mobiles). This way, they will remove / cancel the bank cards. Next, they will be possibly decide to cancel our fingers!
- Resolving Weaknesses of NLPs – Make rule-based NLP scientists to work together and exchange technologies – Share their data resources – Spread the knowledge (and the rules) humans create.
- Resolving Weaknesses of LLMs – Set the rule-based NLPs to correct the weaknesses of LLMs – NLPs in the role of mentor for the LLMs – NLPs auditing the LLMs
- mutually beneficial synergy of the two technologies (κοινή επωφελία αμοιβαίων συνλειτουργιών των δύο τεχνολογιών)

Resolving Weaknesses of LLMs – NLP Auditing 1/4

original	LLM result	correct conversion
in English	comment	in English
<p>Ο Δημήτρης αντιλαμβάνεται τις εξισώσεις γιατί είναι έξυπνος</p> <p>[Ο Dimitris <u>an-</u> <u>tilamvanetai tis exisoseis</u> <u>giati einai exypnos</u>]</p>	<p>Οι εξισώσεις αντιλαμβάνονται από τον Δημήτρη γιατί είναι έξυπνος</p> <p>[Οι <u>exisoseis an-</u> <u>tilamvanontai apo ton</u> <u>Dimitri giati einai ex-</u> <u>ypnos</u>]</p>	<p>Οι εξισώσεις γίνονται αντιληπτές από το Δημήτρη γιατί είναι έξυπνος</p> <p>[Οι <u>exisoseis ginontai antiliptes apo</u> <u>to Dimitri giati einai exypnos</u>]</p>
<p>Dimitris understands equations because he is smart</p>	<p>It is an unnatural / strange sentence</p>	<p>The equations are understood by Dimitris because he is smart</p>

Resolving Weaknesses of LLMs – NLP Auditing 2/4

original	LLM result	correct conversion
in English	comment	in English
Ο Παναγιώτης βαριέται τα πάρτι [Ο Panagiotis <u>varietai ta</u> <u>parti</u>]	Τα πάρτι βαριούνται από τον Παναγιώτη [Τα <u>parti variountai</u> <u>apo ton Panagioti</u>]	Τα πάρτι γίνονται βαρετά στον Πα- ναγιώτη [Τα <u>parti ginontai vareta ston</u> <u>Panagioti</u>]
Panagiotis is bored of parties	It is completely not acceptable by native speakers	Parties are getting boring for Panagiotis

Resolving Weaknesses of LLMs – NLP Auditing 3/4

original	LLM result	correct conversion
in English	comment	in English
Εκμεταλλεύονται τους φυσικούς πόρους [Ekmetallevontai tous fysikous porous]	Οι φυσικοί πόροι εκμεταλλεύονται [Oι <u>fysikoi poroi ekmetallevontai</u>]	Οι φυσικοί πόροι είναι εκμεταλλεύσιμοι OR Οι φυσικοί πόροι γίνονται αντικείμενο εκμετάλλευσης [Oι <u>fysikoi poroi einai ekmetallefsimoi</u> OR Oι <u>fysikoi poroi ginontai antikeimeno ekmetallefsis</u>]
<u>They exploit natural resources</u>	It is an unnatural / strange sentence	Natural resources are exploitable OR Natural resources become object of exploitation

Resolving Weaknesses of LLMs – NLP Auditing 4/4

- Rule-based NLP/NLUs are the solution. There are specific rules for converting sentences from active syntax to passive syntax when the verb of sentence is an ‘αποθετικό’ [apothetiko] (deponent) verb. LLMs are not able (as our examples prove) to learn such rules but Rule-based (or True) NLP/NLUs have this knowledge (Rules existing in the literature which have been encoded to computer programs by humans). We can mention one such (of the available rules):
 - Auxiliary verb (είμαι [eimai] / γίνομαι [ginomai]) + adjective form of the verb in passive
 - Examples are:
 - αντιλαμβάνομαι → γίνομαι αντιληπτός [antilamvanomai → ginomai antiliptos]
 - αντιλαμβάνονται → γίνονται αντιληπτές [antilamvanontai → ginontai antiliptes]
 - (Greek verbs have person and number but Greek adjectives have case, number and sex)
- Consequently, in its simplest way, Rule-based (or True) NLP/NLUs can suggest (send prompts) to LLMs of what auxiliary verb (είμαι / γίνομαι) and which adjectival form of the verb to be used for the passive syntax conversion of a sentence having ‘αποθετικό’ [apothetiko] (deponent) verb.

Benefits and Challenges for Sciences

- Digitize your/their data
- Transform your/their tacit/implicit into explicit knowledge
(σιωπηρή και ρητή γνώση)
- Impose your rules for Auditing LLMs

THANK YOU!

- ANY QUESTIONS

- CONTACT

Nikitas N. Karanikolas

University of West Attica (UniWA)

Department of Informatics and Computer Engineering

nnk@uniwa.gr

<http://users.uniwa.gr/nnk/>

Clarifications

- Some materials in this slides are additional from the original presentation (June 5, 2025).
- Other three works of mine are:
 - “Large Language Models versus Natural Language Understanding and Generation”, PCI’2023 conference, <https://doi.org/10.1145/3635059.3635104>,
 - “Strengths and Weaknesses of LLM-Based and Rule-Based NLP Technologies and Their Potential Synergies”, MDPI Electronics 14(15), 2025, <https://doi.org/10.3390/electronics14153064>,
 - “Understanding and Establishing Natural Language Understanding & Natural Language Generation Technologies for reliable, ethical AI systems that don't impoverish the human languages”, BCI’2025 conference, find the publication of this work following the link <https://users.uniwa.gr/nnk/papers/EstabNLUandNLGTechForSysNotImpoverLang.pdf>.
- Left part of slide 9 (Phrase structure Tree) is same with figure 5 of my paper in PCI’2023.
- Right part of slide 9 (Dependency Tree) is same with a figure of my paper in BCI’2025.
- Phrase structure trees (nested lists) of Slide 10 are same with figure 6 of my paper in PCI’2023.
- Slides 11, 27-30 are from material of my paper in MDPI Electronics.