CLIR Experiments on Perseus Corpus

Shibamouli Lahiri

Recap

- Our short-term goal is to build a multilingual search engine

- The corpus is in Ancient Greek

- A user can query in English, Modern Greek or Ancient Greek

- For now we only focus on English and Ancient Greek queries

Motivation behind our experiment

- No Ancient Greek – English cross-lingual search system is available

- Even translation from Ancient Greek free text to English is not good enough (by Google Translate)

Google Translate does not help

- Herodotus, "The Histories", Book 1, Chapter 1, Section 0

- Ancient Greek: "Ἡροδότου Ἀλικαρνησσέος ἱστορίης ἀπόδεξις ἥδε, ὡς μήτε τὰ γενόμενα ἐξ ἀνθρώπων τῷ χρόνῳ ἐξίτηλα γένηται, μήτε ἔργα μεγάλα τε καὶ θωμαστά, τὰ μὲν Ἕλλησι τὰ δὲ βαρβάροισι ἀποδεχθέντα, ἀκλεᾶ γένηται, τά τε ἄλλα καὶ δι᾽ ἣν αἰτίην ἐπολέμησαν ἀλλήλοισι."

- Human translation (A.D. Godley, Harvard University Press, 1920): "This is the display of the inquiry of Herodotus of Halicarnassus, so that things done by man not be forgotten in time, and that great and marvelous deeds, some displayed by the Hellenes, some by the barbarians, not lose their glory, including among others what was the cause of their waging war on each other."

- Google Translate: "Herodotus Alikarnisseos istoriis apodexis ide, ὡς nor τὰ contemplated ἐξ people unto χρόνῳ genitai faded, nor works great II. THOMAS, as regards the ELLIS And the varvaroisi were accepted, aklea genitai, GFs, but also self aitiin epolemisan not they,."

- Yahoo Translate (AltaVista): "H[rodotoy] A[likarnisseos] i[storiis] ά[podexis] ή[de], ώ[s] [mite] [t]ὰ [genomena] ἐ[x] ά[nthropon] [t]ῷ [chron]ῳ ἐ[xitila] [genitai], [mite] ἔ[rga] big [te] [ka]ὶ [thomasta], [t]ὰ [m]ὲ[n] Ἕ[llisi] [t]ὰ [d]ὲ [barbaroisi] ά[podechthenta], ά[kle]ᾶ [genitai], the [te] ά[lla] [ka]ὶ [di]᾽ ἥ[n] [a]ἰ[tiin] ἐ[polemisan] ά[lliloisi]."
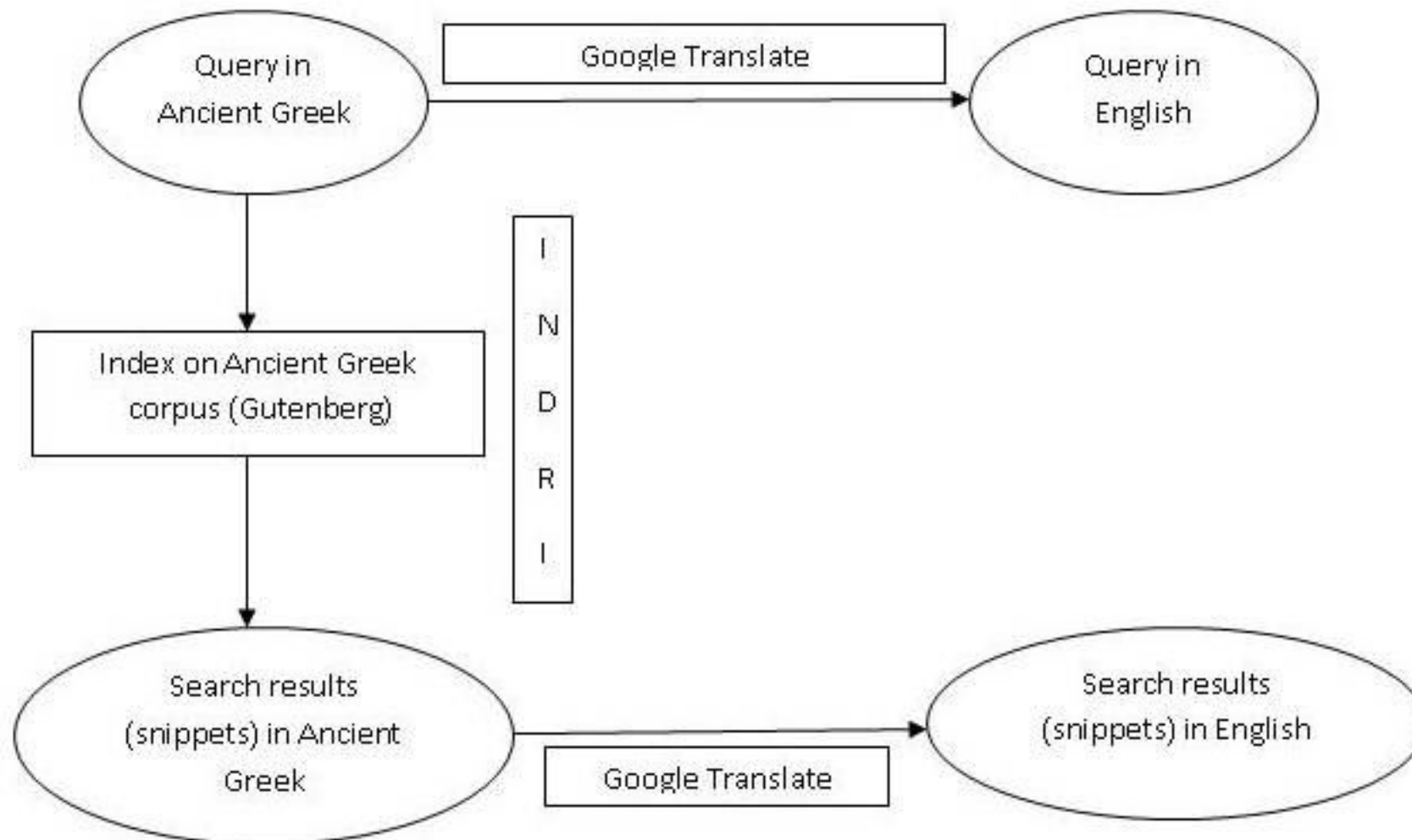
4

An experiment

- To understand the difficulties and challenges involved in Ancient Greek – English CLIR

- Relevance judgment

- Importance of dictionaries in translation process

- Do we need a parallel corpus?

Setup

- Perseus corpus – Ancient Greek texts with their English translation (87 books)

- Gutenberg corpus: we only used Ancient Greek texts (39 books)

- Lexilogos – An Ancient Greek ←→ English dictionary

- Lemur (Indri) search engine

First Experiment

Query in Ancient Greek → Google Translate → Query in English

Query in Ancient Greek → Index on Ancient Greek corpus (Gutenberg) → Search results (snippets) in Ancient Greek

INDRI

Search results (snippets) in Ancient Greek → Google Translate → Search results (snippets) in English

Results

- **Query:** αρσενικός (male, masculine)

- **Snippet 1:** ...και μόνη, σε λιγάκι. Να μη βρεθή **αρσενικός**, που δόρυ να σηκώση, και άλλον να... (...and only a little. Do not get me **male**, who can lift a spear, and else...)

- **Snippet 2:** ...Ότι όμως ήσουν μάντις και ότι υπήρξες **αρσενικός** και θηλυκός το γνωρίζω από τους ποιητάς... (...But that you were a seer, and that was **male** and female I know of poets...)

- **Snippet 3:** ...64} Κανείς εκ των θεών, θηλυκός ή **αρσενικός**, ας μη λείψη ούτε εκ των ποταμών... (...64} None of the gods, female The **male**, let's not miss either of rivers...)
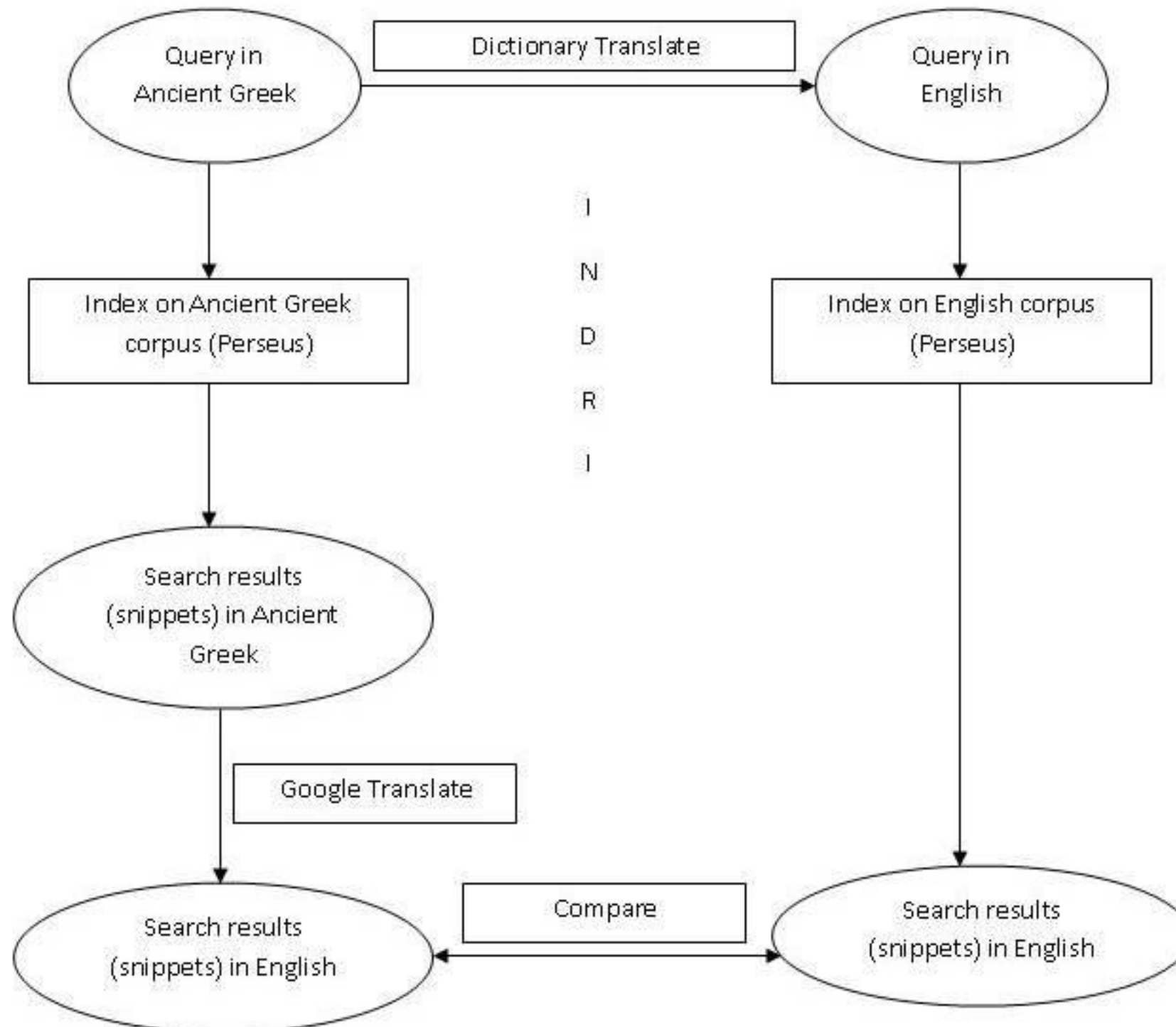
Results (Contd.)

- **Query:** φαίνεται (seems)

- **Snippet 1:** …απομείνουν τα ευλογοφανή, τότε είναι αρκετή απόδειξις τούτο. ΚΕΦΑΛΑΙΟΝ Β'. &Προβλήματα εγκρατείας και ακρατείας&. — **Φαίνεται** δε ότι η εγκράτεια και η καρτερία ανήκει εις τα σπουδαία και αξιέπαινα, η δε…Διότι, επειδή θέλει να εξελέγξη ως ψευδή πράγματα τα οποία δεν είναι ευλογοφανή, διά να **φαίνεται** δεινός όταν επιτύχη, κάμνει συλλογισμόν ο οποίος αποτελεί απορίαν. Δηλαδή δεσμεύεται η διάνοια, όταν δεν…

- (…the remaining plausible, then it is quite demonstrable purpose. CHAPTER II. & Problems continence and incontinence &. - It **seems** that abstinence and endurance is at the important and laudable, and … Because, because he wants to auditing as false things that are not is presented, by look awesome when you succeed, sick reasoning which is aporian. That is bound the intellect, when no…)

- **Snippet 2:** …σαφέστατα το κακόστομον επίρρημα &κατεσπευσμένως,& το οποίον συγκρινόμενον με τα γοργά ευρωπαϊκά vite, schnell κτλ., **φαίνεται** ως ειρωνεία προς την αλήθειαν (=μη κουνιέσαι) και το οποίον ημείς το λέγομεν γλήγορα εις…ως σκοπός εις όλα&. — Πάσα τέχνη και επιστήμη, καθώς και πάσα πράξις και προαίρεσις, **φαίνεται** ότι επιδιώκει κάποιον αγαθόν. Δια τούτο καλώς απεφάνθησαν ότι αγαθόν είναι ό,τι επιδιώκουν πάντες…

- (…clearly the kakostomon adverb Katespefsmenos &, & that compared with the fast-European vite, schnell, etc., would **seem** ironic to the truth (= non-cot) and we ourselves why we say the gligora … as to the purpose at all &. - Pass art and science, and and any proairesis and Praxis, **appears** to seek a goods. Wherefore rightly apefanthisan that good is what seeking everybody…)

Observations

- Looks like Google Translate is doing a decent job of translating snippets.

- Let's see what happens for Perseus corpus.

Second Experiment

Query in Ancient Greek → Dictionary Translate → Query in English

Query in Ancient Greek → Index on Ancient Greek corpus (Perseus)

Query in English → Index on English corpus (Perseus)

I N D R I

Index on Ancient Greek corpus (Perseus) → Search results (snippets) in Ancient Greek

Search results (snippets) in Ancient Greek → Google Translate → Search results (snippets) in English

Index on English corpus (Perseus) → Search results (snippets) in English

Search results (snippets) in English → Compare → Search results (snippets) in English

Results

- Query: λεχέων (couch, bed, bedstead)

- Snippet 1:   ...ᾱ̃τερ φιλότητοσ ἐφιμέρου, οὐδέ οἱ ἦεν πρὶν **λεχέων** ἐπιβῆναι ἐυσφύρου Ἠλεκτρυώνης, πρίν γε φόνον τίσαιτο... (...ater filotitos ἐφιμέρου, οὐδέ οἱ before or in **λεχέων** epivinai ἐυσφύρου Ἠλεκτρυώνης, πρίν GI murder τίσαιτο...)

- Snippet 2:   ...ᾱ̃ θ' Ἕλλανα λιποῦσα δόμον Φρυγίων **λεχέων** ἔπλευσε πλαθεῖσ' ὑπ' Ἰλίῳ &δαγγερ... (...a i–FAT Ellanios domon Φρυγίων **λεχέων** sailed platheis' In Ἰλίῳ & danger...)

- Snippet 3:   ...μήποτε μήποτέ μ', ὦ πότνιαι Μοῖραι, **λεχέων** Διὸσ εὐνά-τειραν ἴδοισθε πέλουσαν: μηδὲ πλαθείην...   (...μήποτε μήποτέ me, O potniai Inevitably, **λεχέων** DIOS-εὐνά teiran idoisthe πέλουσαν: zero πλαθείην...)

- **Google Translate did not work!**

English sense 1

- **Query:**   couch

- **Snippet 1:**   ...of Trojan guards. Chorus Go to Hector's COUCH. Which of you squires that tend the...eye from its repose; quit your lowly COUCH of scattered leaves, Hector! It is time to hearken. Hector Who is this? Is it a friend who calls? Who are you? Your password? Speak! Who are these who come near my COUCH in the night? You must tell me...

- **Snippet 2:**   ...Metaneira, and she rose up from her COUCH before Demeter, and bade her be seated. But Demeter, bringer of seasons and giver of perfect gifts, would not sit upon the bright COUCH, but stayed silent with lovely eyes cast...Hades in his house seated upon a COUCH, and his shy mate with him, much...

- **Snippet 3:**   ...him lay the lady his wife, who had strewn the COUCH. Soon as early Dawn appeared, the rosy-fingered,up from...the hall with torches in their hands and strewed the COUCH, and a herald led forth the guests. So they slept...thought, any one of you,to rouse me from my COUCH, though in your hearts ye knew full well when he...

- **The parallel English documents show better snippets.**

English sense 2

- **Query:**   bed

- **Snippet 1:**   ...purpose. And whenever I make here my BED, restless and dank with dew and unvisited...aloud the signal to rise from her BED, and as quickly as she can to...of their wings, wheel high over their BED, because they have lost the toil of...for the princes! Alas for the husband's BED and the impress of her form so...

- **Snippet 2:**   ...should not pity your country or my BED, for this my marriage will destroy those...on the ground instead of my royal BED lay down my shrunken limbs, with tattered...but now am a slave, to a BED upon the ground, near some rocky ridge...Phrygian blood, and desolation reigned over every BED where young men lay beheaded, a glorious...

- **Snippet 3:**   ...can have no better advice than to keep warm and comfortable in BED, to take a purge, and to adopt a diet of slops. Within...may perhaps add the treatise Decorum, which deals (among other things) with BED-side manners, and Precepts, a work similar in style and subject. The...

English sense 3

- **Query:** bedstead

- **Snippet 1:** ...on a peg beside the corded1 BEDSTEAD, and then went forth from the chamber...Odysseus, to sleep there on a corded BEDSTEAD under the echoing portico,and by him...armed Arete bade her maidens place a BEDSTEAD under cover of the portico, and to...they had busily spread the stout-built BEDSTEAD, they came to Odysseus, and called to...

- **Snippet 2:** ...I saw her spread coverings on the BEDSTEAD of Heracles. When she had done this...

- **Snippet 3:** ...With this he groped about for the BEDSTEAD, and sitting down by my feet he...
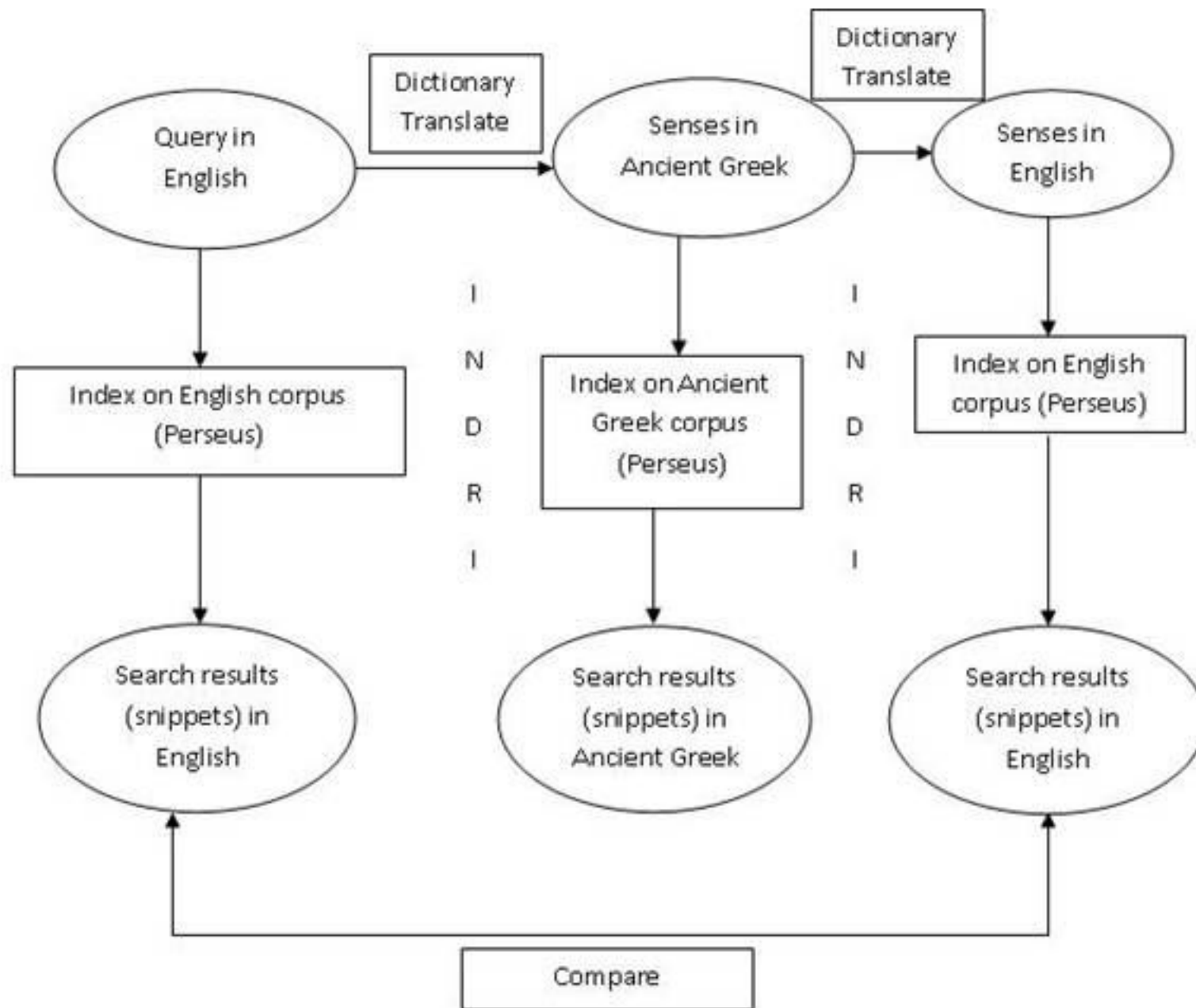
Relevance

- We compared top 5 outputs returned by the original Greek query and the three translated English senses.

Greek (λεχέων)	English 1 (couch)	English 2 (bed)	English 3 (bedstead)
hes.sh	eur.rh	aesch.ag	hom.od
eur.rh	hh	eur.tro	soph.trach
aesch.pb	hom.od	hp.jones	plat.tet6
hh	soph.ot	Document ID	n.il
eur.tro	hom.il	hes.th	

Third Experiment

Results

- **Query:** troy

- **Snippet 1:** ...that Phoebus and I with exact measurement set towers of stone about this land of TROY and ringed it round, never from my heart has passed away a kindly feeling for...Zeus, or from some lesser power? Athena From none of these; but on behalf of TROY, whose soil we tread, I have come to seek your mighty aid, to make it...

- **Snippet 2:** ...hands in days of old built the walls of TROY. Chorus May he come to the ships! May he...has come against my city, against the land of TROY with his army of a thousand ships. Messenger Lord...Did you say that Rhesus was setting foot in TROY? Messenger You have it; and lighten me of half...

- **Snippet 3:** ...from mistrust and danger, while mine cannot. You have sailed to TROY under no oath to any man, nor under any constraint. Neither...what gain is it for me that he should come to TROY? Odysseus His arrows alone will capture TROY. Neoptolemus Then I am not to be the conqueror, as you...

After query translation

- **Query:** Ἴλιος

- **Snippet 1:** ...σπανίζομεν. Ἕκτωρ τί δῆτα χρῄζεισ ὧν κέκευθεν **Ἴλιος**; Δόλων ἑλὼν Ἀχαιοὺσ δῶρά μοι ξυναίνεσον. Ἕκτωρ...

- **Snippet 2:** ...ἔτ' ἔστι Τροία. Ἑκάβη ὀττοτοτοτοτοῖ. λέλαμπεν **Ἴλιος**, Περ- γάμων τε πυρὶ καταίθεται τέραμνα καὶ...

- **Snippet 3:** ...ὁ Ἕλενοσ λέγει πῶσ ἂν αἱρεθείη ἡ **Ἴλιος**,...

After query translation (Contd.)

- **Query:**   Ἴλιον

- **Snippet 1:**   ...', αἶ συνεξεῖλον Φρύγας λείπω τὸ κλεινὸν Ἴλιον βωμούσ τ' ἐμούς: ἐρημία γὰρ πόλιν...ὕπο. Ποσειδῶν καὶ μὴν ἔπερσάν γ' Ἴλιον τῷ σῷ σθένει. Ἀθήνα τοιγάρ σφε σὺν...ἅτασ κελαδεῖν ἀχορεύτους. Ἑκάβη πρῷραι ναῶν, ὠκείαισ Ἴλιον ἱερὰν αἶ κώπαισ δι' ἅλα πορφυροειδέα...ἅλοχοι μέλεαι, καὶ κοῦραι κοῦραι δύσνυμφοι, τύφεται Ἴλιον, αἰάζωμεν. μάτηρ δ' ὡσεί τισ πτανοῖσ...

- **Snippet 2:**   ...ἐόντα τά τ' ἐσσόμενα πρό τ' ἐόντα, καὶ νήεσσ' ἡγήσατ' Ἀχαιῶν Ἴλιον εἴσω ἦν διὰ μαντοσύνην, τήν οἱ πόρε Φοῖβοσ Ἀπόλλων: ὅ σφιν ἐὺ φρονέων ἀγορήσατο καὶ...Ζεύσ με μέγα Κρονίδησ ἅτῃ ἐνέδησε βαρείῃ, σχέτλιος, ὃσ πρὶν μέν μοι ὑπέσχετο καὶ κατένευσεν Ἴλιον ἐκπέρσαντ' εὐτείχεον ἀπονέεσθαι, νῦν δὲ κακὴν ἀπάτην βουλεύσατο, καί με κελεύει δυσκλέα Ἄργοσ ἱκέσθαι...

- **Snippet 3:**   ...θέλοντασ ὠφελεῖν πόλιν. Ἕκτωρ ἀρκοῦμεν οἱ σῴζοντεσ Ἴλιον πάλαι. Χορόσ πέποιθασ ἤδη πολεμίουσ ᾐρηκέναι; Ἕκτωρ...μοι, Σκύθησ λεώς, μέλλοντι νόστον τὸν πρὸσ Ἴλιον περᾶν ξυνῆψε πόλεμον: Εὐξένου δ' ἀφικόμην...' ἦν ὡς ἔπλευσ' ἐπ' Ἴλιον. Ἕκτωρ ἔπλευσε καὶ πάρεστιν: ἀλλὰ μηνίων στρατηλάταισιν...

Translating back to English

- **Query:**   Ilion

- **Snippet 1:**   ...him, you made Troy, the city of ILION, to be no city. He did not...

- **Snippet 2:**   ...them bring him at once to windy ILION. And if any mortal man ask you...

- **Snippet 3:**   ...the sea, just as the country about ILION and Teuthrania and Ephesus and the plain...the Cyprian poems relate that Alexandrus reached ILION with Helen in three days from Sparta...disposing their forces, they sent messengers to ILION, one of whom was Menelaus himself. When...I reason thus: had Helen been in ILION, then with or without the will of...

Relevance

- We compared top 5 outputs returned by the original English query, the two translated Greek queries and the final English query obtained by translating the Greek queries.

English (troy)	Greek 1 (Ἴλιος)	Greek 2 (Ἴλιον)	Back to English (Ilion)
eur.tro	eur.rh	eur.tro	aesch.eum
eur.rh	eur.tro	hom.il	hh
soph.phil	apollod	eur.rh	hdt
eur.ia		soph.phil	apollod
apollod		eur.ia	plat.rep

Overall

- We need **query**, snippet and **document** translation.

- **Google Translate** does not help much in Ancient Greek snippet or query translation.

- For query translation, **dictionaries** are right now the best.

- **Morphological analysis** needed (Ancient Greek is highly inflected)

- Snippet and Document translation requires statistical machine translation (**SMT**) approach – at sentence or word level

- For relevance judgment, **evaluation** needs to be done on a number of different queries and different translations

Help needed!

- Data

- Tagged data

- Evaluation of search result relevance

- Evaluation of query, snippet and document translation

Value of tagged data

- A mini parallel corpus, if the tags are in English

- Queries based on tags can be answered easily

- A new query can be matched with tags to obtain better search results

- Tags provide topics; a topic-based search can therefore be implemented based on tags

- Translation evaluation is eased once we have tagged data

Working on

- **Put up a user interface on the web for searching Ancient Greek documents**

Working on (Contd.)

- Translation – query and snippet for now, documents later

- Methods – dictionary, SMT

CLIR Experiments on Perseus Corpus

Shibamouli Lahiri