

Computational Linguistics 1

Radhika Mamidi

Warm-Up Exercise

Translate the following sentences into your mother tongue and analyse

- I love rainy season. I walk in rain whenever possible. My mother thinks I am mad.
- He is my brother. He hit the ball. The ball hit the window. The window broke.
- The boy who is standing near the gate is my brother.

Tutorial

Discuss your translations for:

- TAM – Tense, Aspect, Modality
- GNP – Gender, Number, Person
- Case
- Active, Passive, Causative etc

Linguistics

Some questions...

- What is language?
- Why do we need language?
- How did language develop?
 - Thousands of languages exist. How?
 - Many are dying. Why?
- Convergence
- Divergence

language, dialect, style,
register, jargon, argot,
cant....

Language families

- Indo-European
- Dravidian
- Sino-Tibetan
- Afro-Asiatic
-

Similarities between languages
belonging to different language families

Genealogical relation or Geographical
proximity?

Writing systems

- Pictographic
- Ideographic
- Syllabic
- Alphabetic
-

What is Linguistics?

- The scientific study of human language
- Aim is to establish a theory by studying nature of a language and by applying this established theory to describe other languages
- Descriptive study not prescriptive
- Speech activity - speaker, hearer, setting, code, medium; plus topic, contextual knowledge, abiding social norms, communicative competence

Applied Linguistics

Subfields / Inter-disciplinary fields:

Sociolinguistics

Lexicography

Semiotics

Stylistics

Computational Linguistics

Psycholinguistics

Anthropological Linguistics

Language Teaching/Learning

Translation Studies

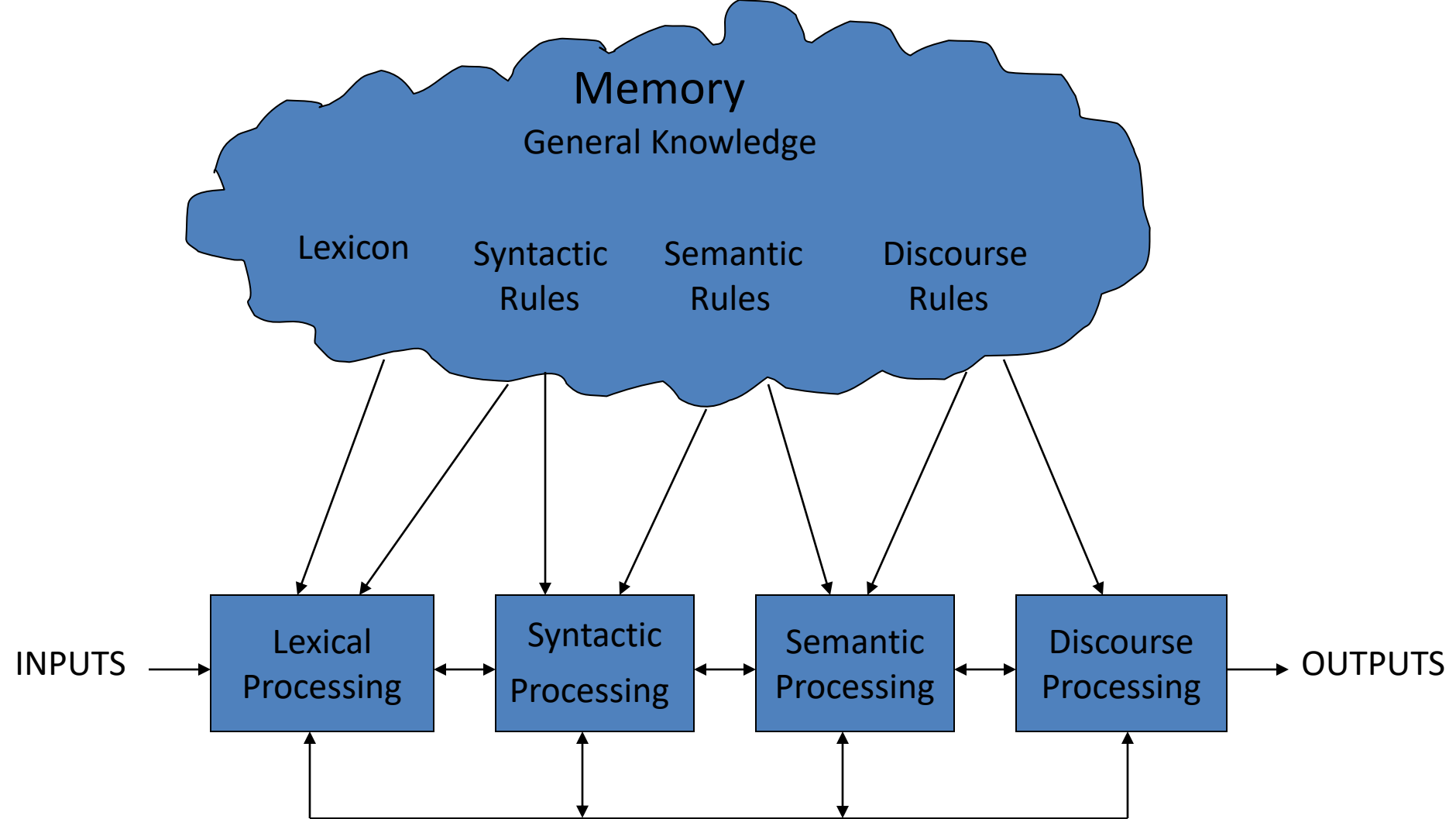
Note: For some, 'Applied Linguistics' is associated with Language teaching (eg ESL, EFL).

Core Linguistics

Levels of Language Processing

- Phonology
 - Morphology
 - Syntax
 - Discourse Analysis
-
- Semantics
 - Pragmatics

- Sound
 - Word
 - Sentence
 - Discourse
-
- Meaning without context
 - Meaning with context



Model of Language Processing

To derive meaning you need all kinds of rules – ‘building’, ‘blocks’

Eg: The building blocks are made of plastic.

The building blocks the sun.

Why is understanding language by a machine so difficult?

- Human language is:
 - Complex and Ambiguous
- We use language creatively
 - We don't mean what we say!
- Language Understanding needs contextual and general knowledge apart from linguistic knowledge.
 - To know what we mean shared knowledge is necessary.

Representing all this knowledge computationally is THE challenge.

Computational Linguistics

Goal of Computational linguistics

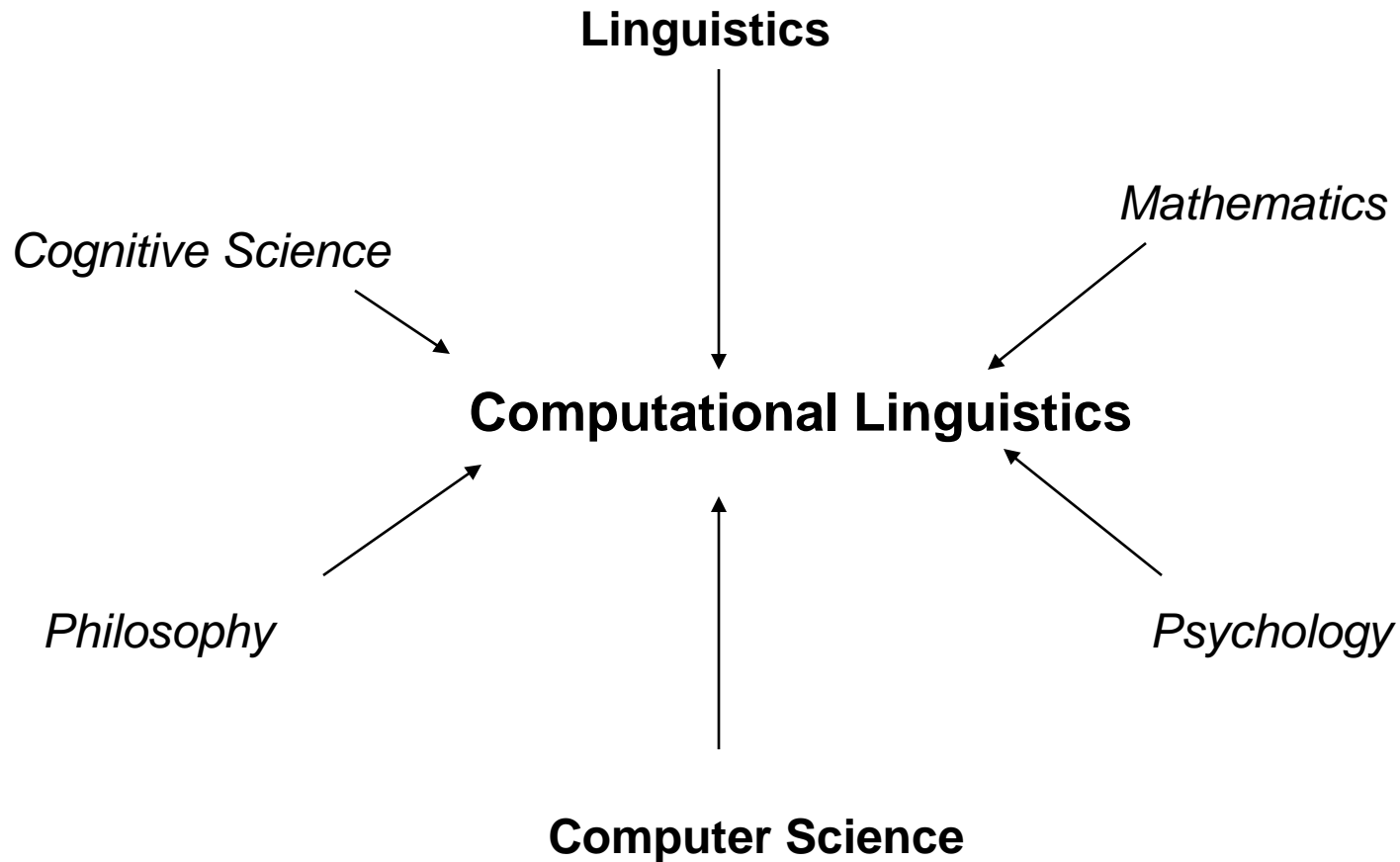
- To make the computers understand human language.
- To use technology to build language tools.
- Use computers to process or produce human language.

We want to communicate with computers in the language we speak!

Why Computers should talk with us?

- The main obstacle in the interaction between human and computer is a communication problem.
- Computers do not understand our language and learning computer language is difficult.
- Computer language do not match the structure of human thought.
- The use of human language in computer applications will benefit the user.

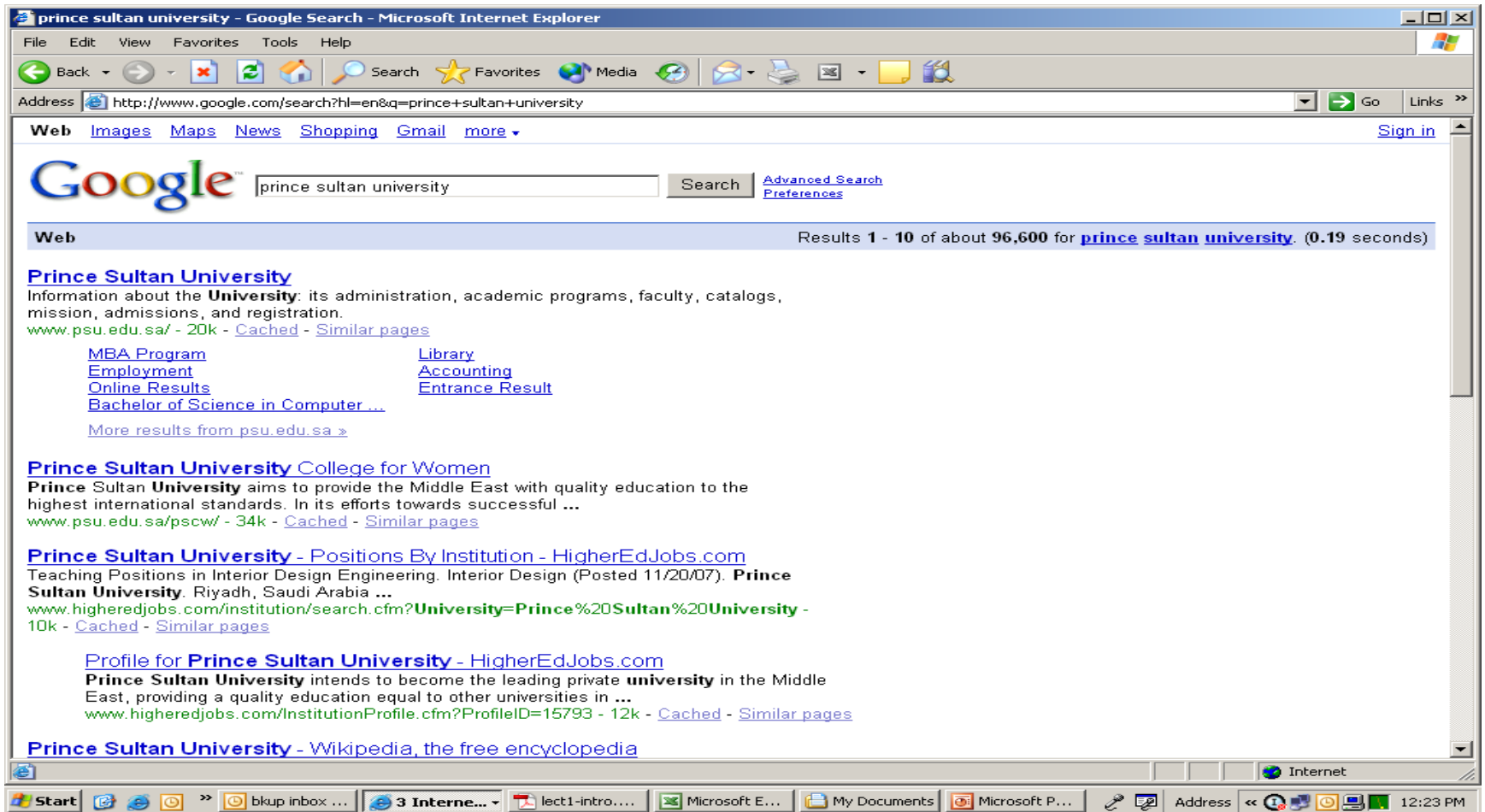
Computational linguistics is an interdisciplinary field. It takes input from different fields.



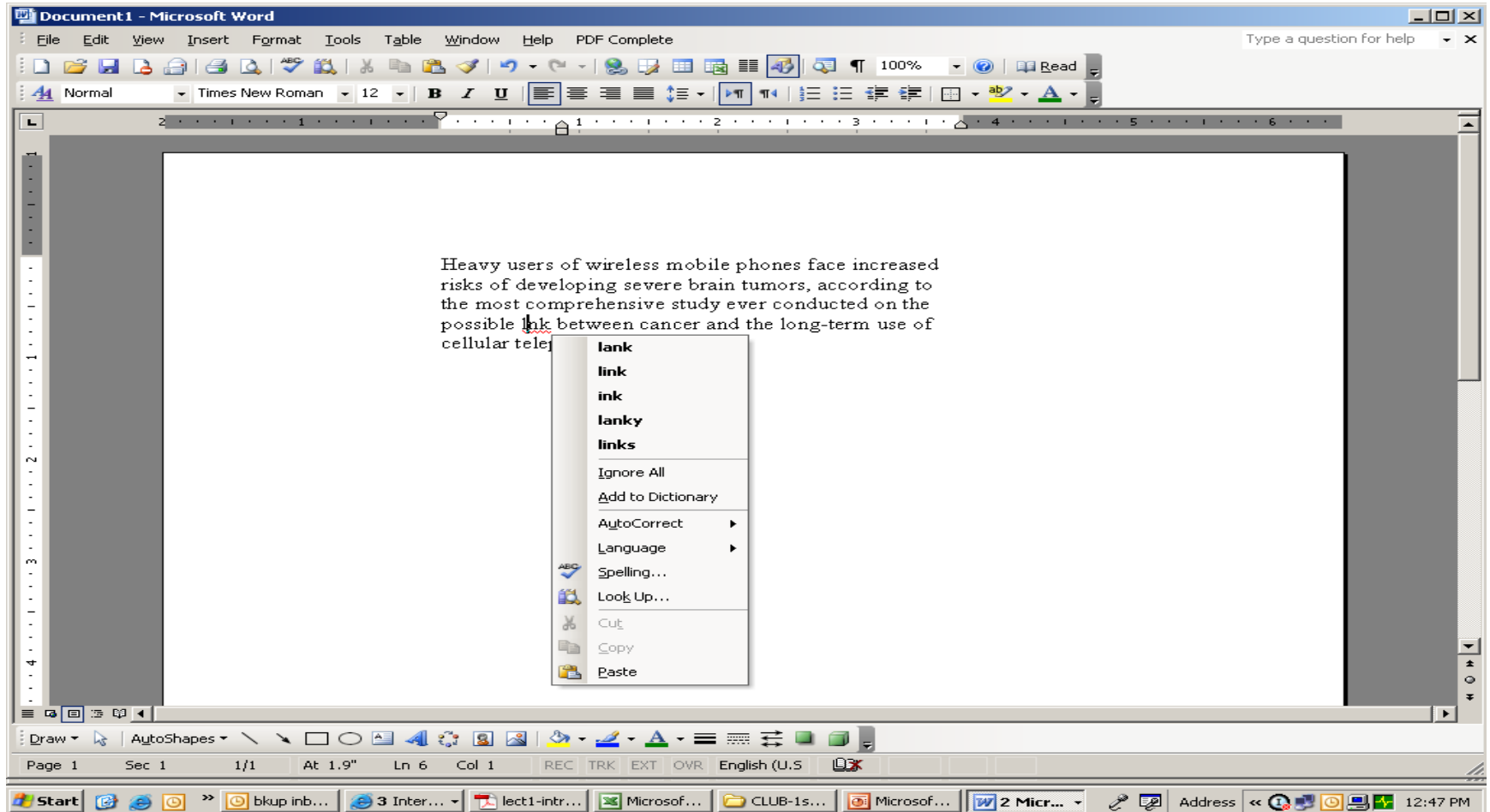
Some Applications of Computational Linguistics

- Spelling and Grammar Checkers
- Screen readers for the blind
- Machine readable Dictionaries
- Machine Translation Systems
- Question-Answering Systems
- Dialogue Systems
- Sentiment Analysers
- Search Engines
- Automatic Summarization

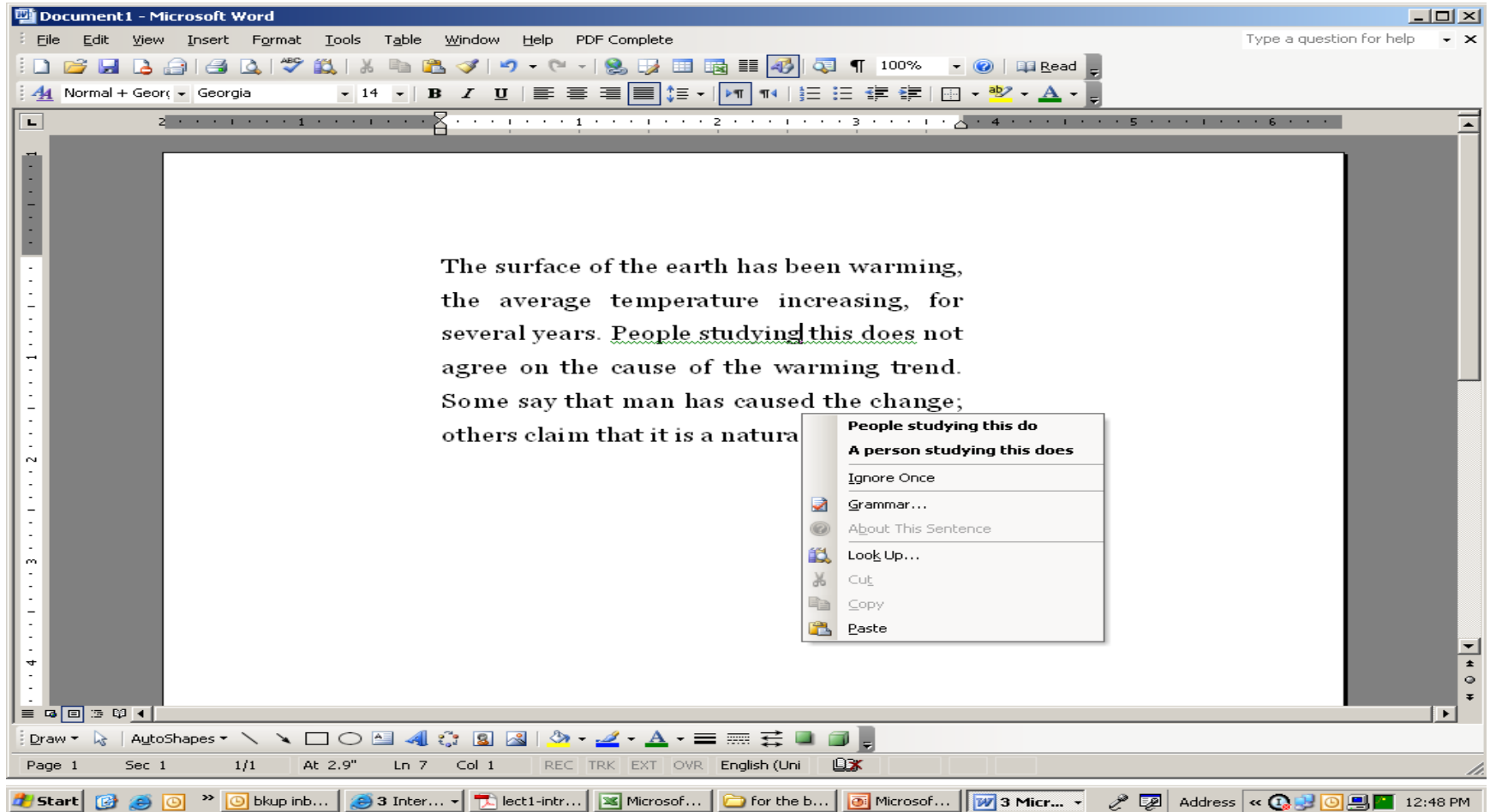
Example Application- Search Engine



Example Application- Spelling Checker



Example Application- Grammar Checker



Example Application- Question Answering

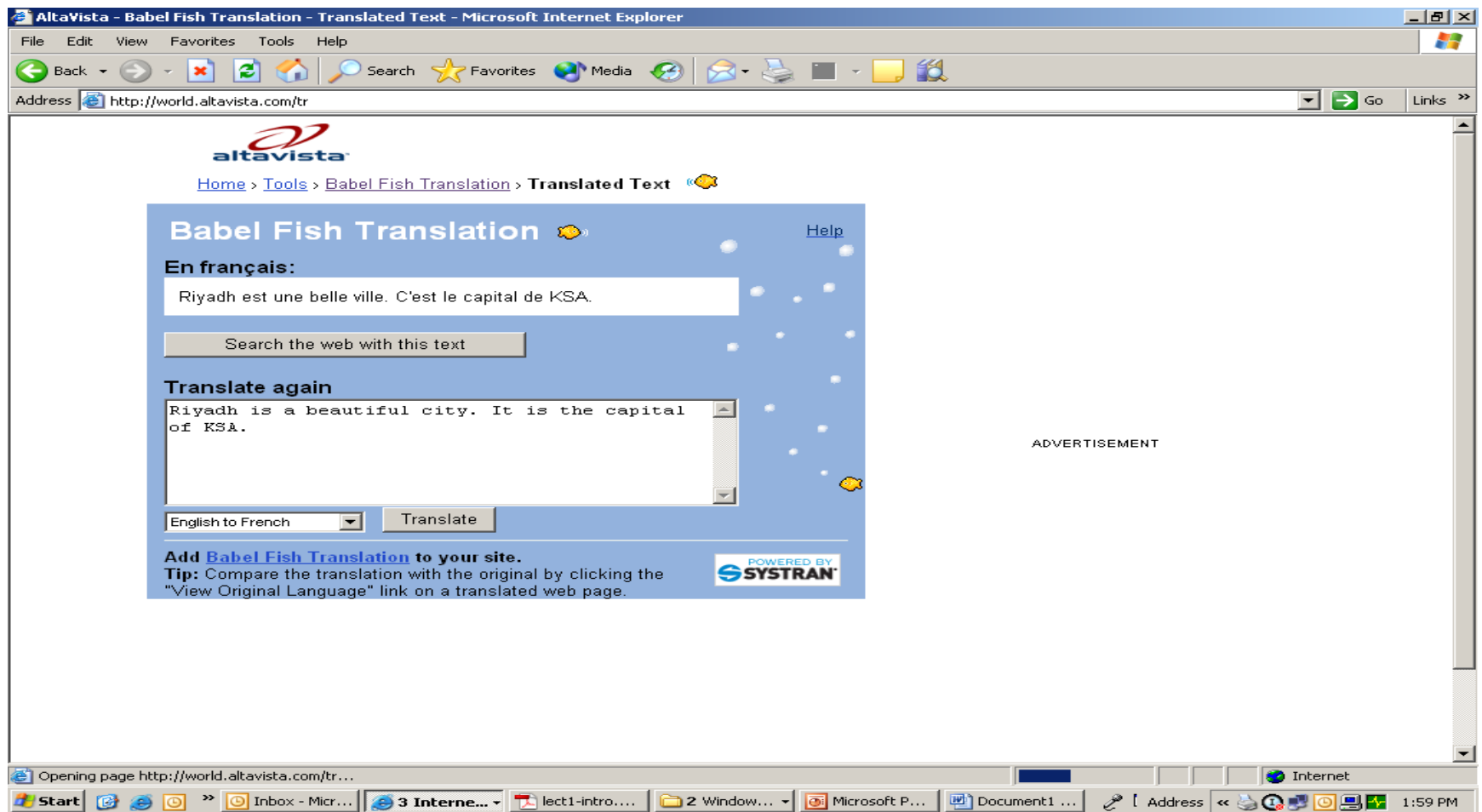
The screenshot shows a Microsoft Internet Explorer browser window with the title "What is the capital of Japan? - Ask.com Web Search - Microsoft Internet Explorer". The address bar shows the URL "http://www.ask.com/web?qsrc=2417&o=0&l=dir&q=What+is+the+capital+of+Japan%3F&search=".

The search results page for "What is the capital of Japan?" is displayed. The main result states: "The capital of **Japan** is Tokyo". It includes links to "World Factbook", "Encyclopedia", "BBC Profile", "US Government Travel Info", and "Maps".

Below the main result, there are sponsored results from "invest-acme99.com" and "invest-clover99.com" offering investment opportunities. Other results include "Japan Geography" from Japan-guide.com, "Capital of Japan - Wikipedia, the free encyclopedia" (stating Tokyo is the capital), "CIA - The World Factbook -- JAPAN -- Government" (providing geographic coordinates and time difference), and "Crowded Commuter Train, Tokyo - Picture - MSN Encarta" (stating Tokyo is the capital and a hub for the railway system).

The left sidebar contains a search bar with the query "What is the capital of Japan?" and a list of related search suggestions under "Narrow Your Search" and "Expand Your Search".

Example Application- Machine Translation



Exercise

- Give your sentences to Google Translate and compare the translations.
- Give your translated sentences to an MT and see the English output. Compare it with your original English input.