NLP Tasks (Continued)

CMSC 473/673 - NATURAL LANGUAGE PROCESSING

Slides modified from Dr. Frank Ferraro & Dr. Jason Eisner

Learning Objectives

Distinguish between different text classification task types

Formalize NLP Tasks at a high-level:

- What are the input/output for a particular task?
- What might the features be?
- What types of applications could the task be used for?

Text Annotation Tasks ("Classification" Tasks)

1. Classify the entire document ("text categorization")
2. Classify word tokens individually
3. Classify word tokens in a sequence
4. Identify phrases ("chunking")
5. Syntactic annotation (parsing)
6. Semantic annotation
7. Text generation

Slide courtesy Jason Eisner, with mild edits

Review: Document Classification

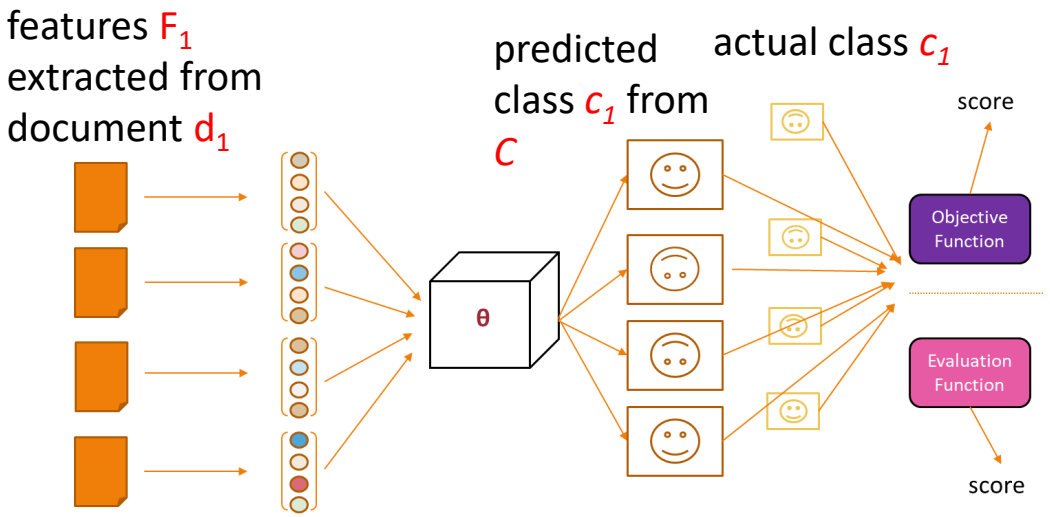
Assigning subject categories, topics, or genres

Spam detection

Authorship identification

Language Identification

Sentiment analysis



What are some example sub tasks or applications?

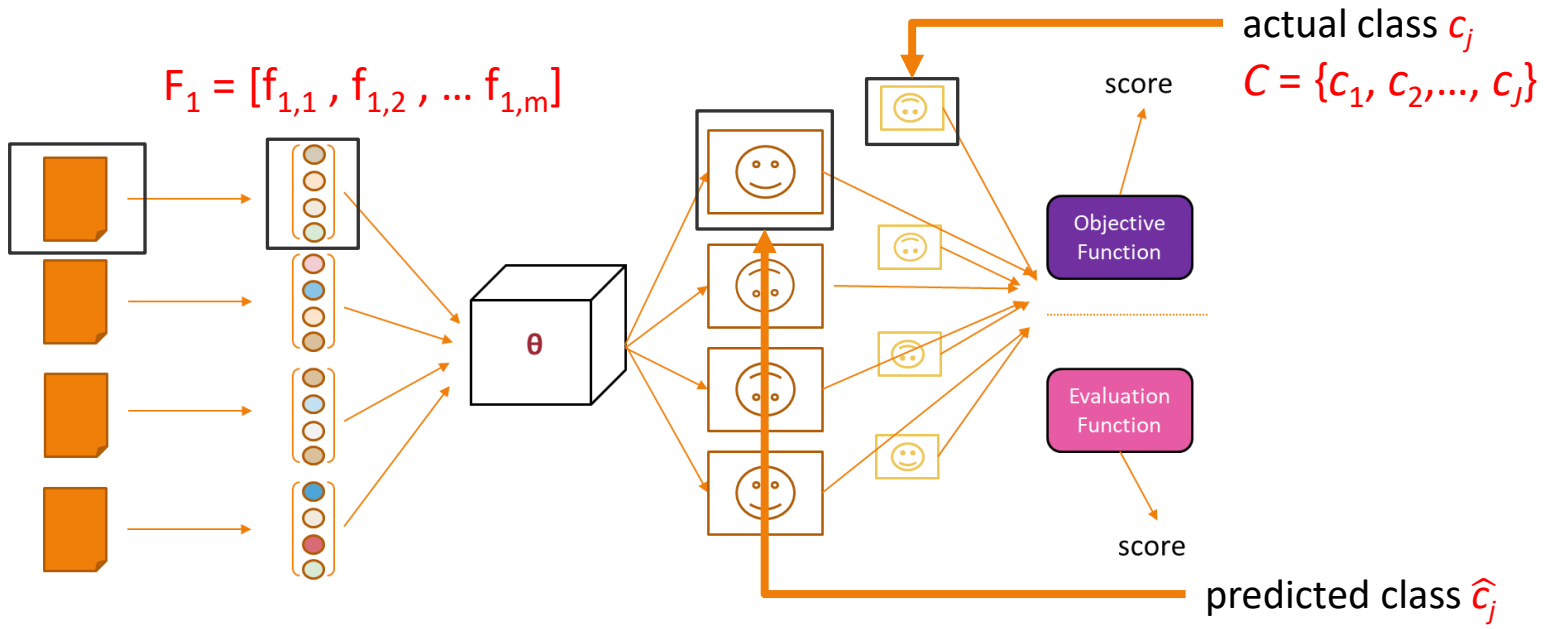
Review: Token Classification

Word pronunciation

Word sense disambiguation (WSD) within or across languages

Accent restoration

features F_1 extracted from word w_1 and its surrounding words (context)

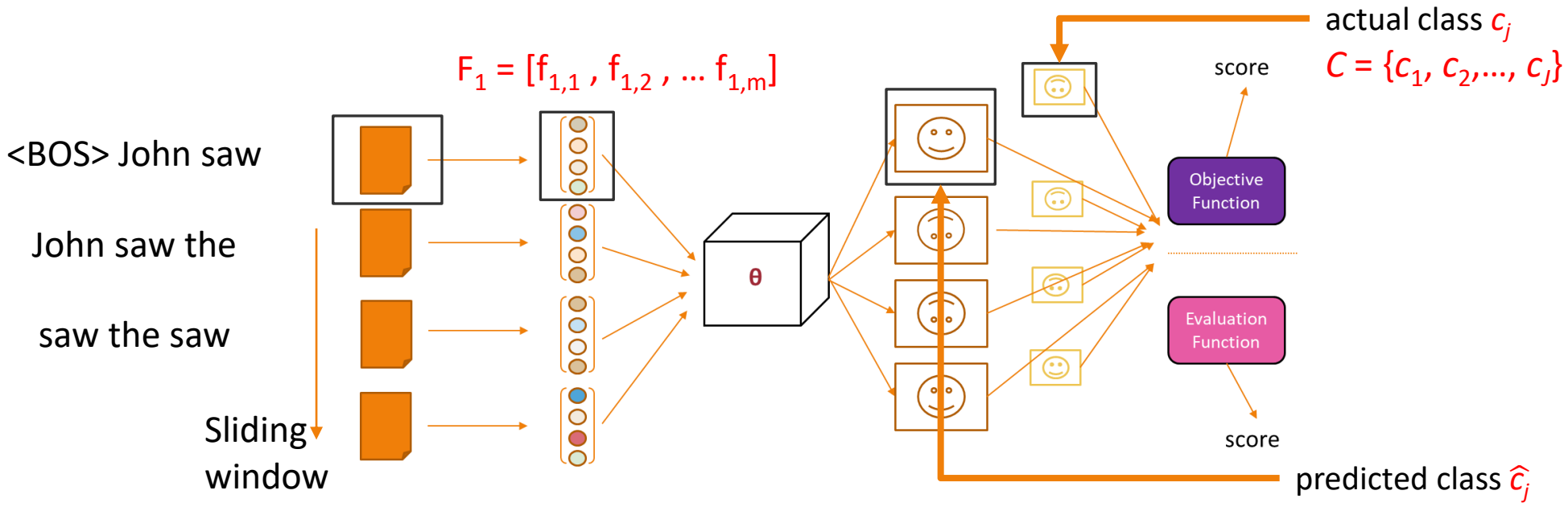


What are some example sub tasks or applications?

Review: Token Classification in a Sequence

<BOS> John saw the saw and decided to take it to the table .

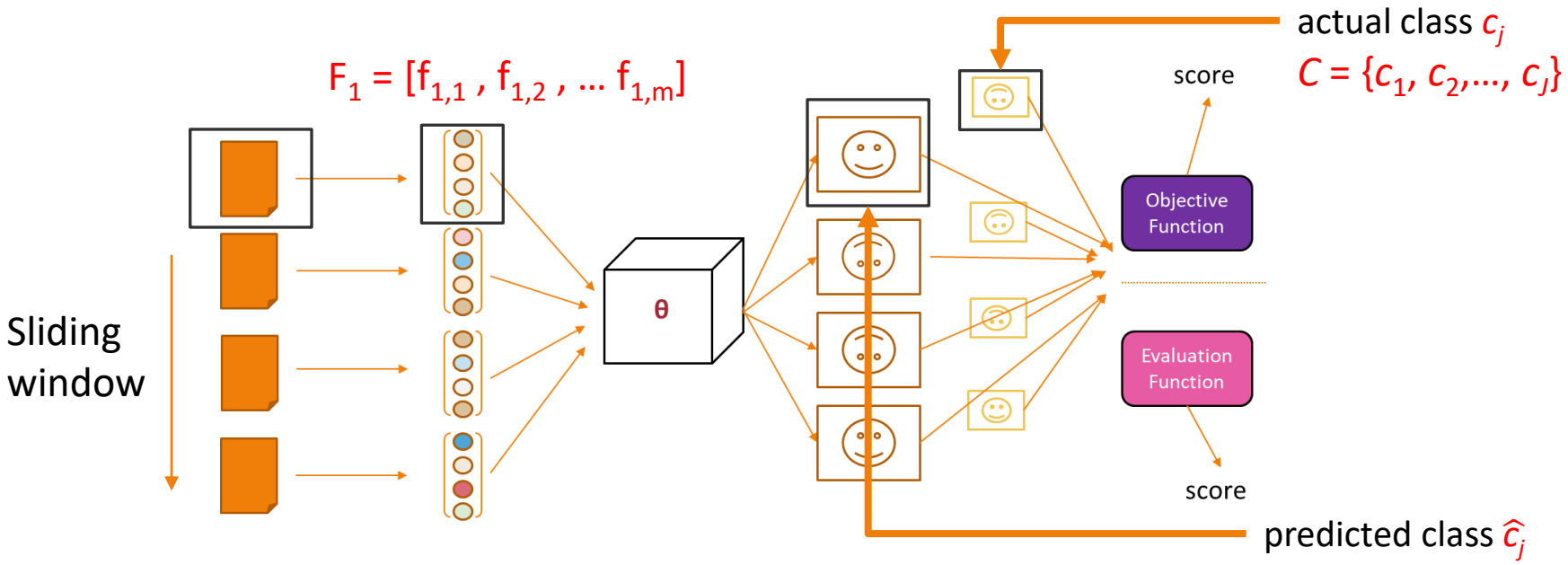
NNP VBD DT NN CC VBD TO VB PRP IN DT NN PUNCT



Review: Token Classification in a Sequence

<BOS> John saw the saw and decided to take it to the table .

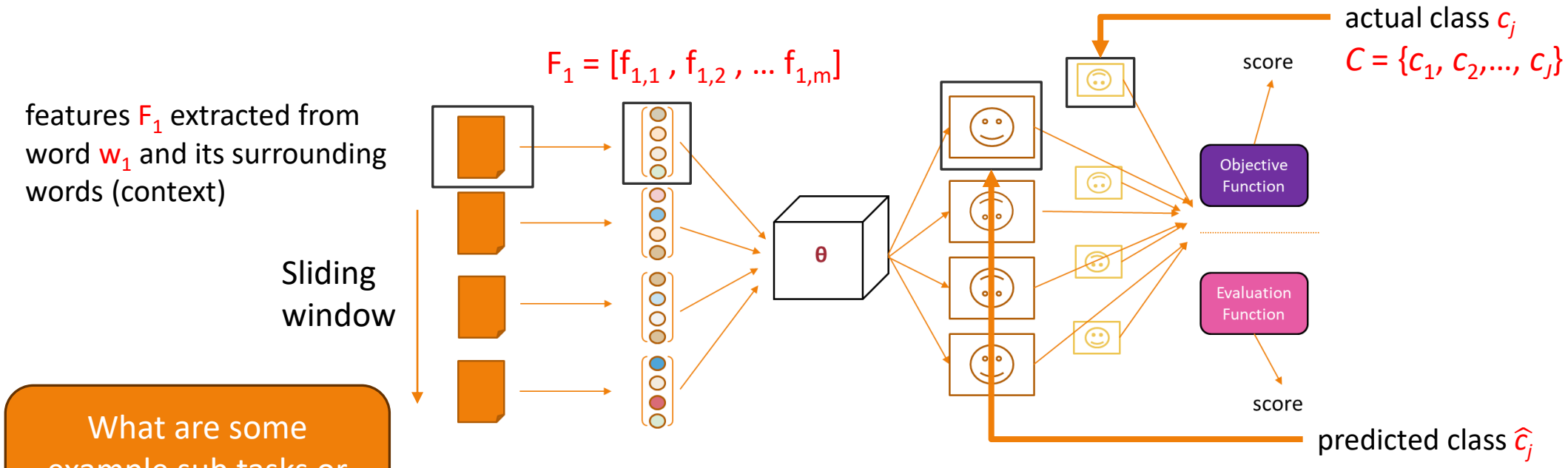
NNP VBD DT NN CC VBD TO VB PRP IN DT NN PUNCT



Review: Token Classification in a Sequence

Part of speech tagging

Word alignment



What are some example sub tasks or applications?

Text Annotation Tasks ("Classification" Tasks)

1. Classify the entire document ("text categorization")
2. Classify word tokens individually
3. Classify word tokens in a sequence
4. Identify phrases ("chunking")
5. Syntactic annotation (parsing)
6. Semantic annotation
7. Text generation

Slide courtesy Jason Eisner, with mild edits

Example: Finding Named Entities

Named entity recognition (NER)

Identify proper names in texts, and classification into a set of predefined categories of interest

- Person names
- Organizations (companies, government organisations, committees, etc.)
- Locations (cities, countries, rivers, etc.)
- Date and time expressions
- Measures (percent, money, weight, etc.),
- email addresses, web addresses, street addresses, etc.
- Domain-specific: names of drugs, medical conditions,
- names of ships, bibliographic references etc.

Cunningham and Bontcheva (2003, RANLP Tutorial)

NE Types

TYPE	DESCRIPTION
PERSON	People, including fictional
NORP	Nationalities or religious or political groups
FACILITY	Buildings, airports, highways, bridges, etc
ORG	Companies, agencies, institutions, etc
GPE	Countries, cities, states
LOC	Non-GPE locations, mountain ranges, bodies of water
PRODUCT	Objects, vehicles, foods, etc (Not services)
EVENT	Named hurricanes, battles, wars, sports events, etc
WORK_OF_ART	Titles of books, songs, etc
LAW	Named documents made into laws
LANGUAGE	Any named language
DATE	Absolute or relative dates or periods.
TIME	Times smaller than a day
PERCENT	Percentage, including "%".
MONEY	Monetary values, including unit
QUANTITY	Measurements, as of weight or distance
ORDINAL	"first", "second", etc
CARDINAL	Numerals that do not fall under another type

https://medium.com/@rajat.jain1/natural-language-extraction-using-spacy-on-a-set-of-novels-88b159d68686

Named Entity Recognition

CHICAGO (AP) — Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

What are the inputs/outputs?

Slide courtesy Jim Martin

Example Use: Information Extraction

As a task: **Filling slots in a database from sub-segments of text.**

October 14, 2002, 4:00 a.m. PT

For years, Microsoft Corporation CEO Bill Gates railed against the economic philosophy of open-source software with Orwellian fervor, denouncing its communal licensing as a "cancer" that stifled technological innovation.

Today, Microsoft claims to "love" the open-source concept, by which software code is made public to encourage improvement and development by outside programmers. Gates himself says Microsoft will gladly disclose its crown jewels--the coveted code behind the Windows operating system--to select customers.

"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access."

Richard Stallman, founder of the Free Software Foundation, countered saying…

IE

NAME	TITLE	ORGANIZATION
Bill Gates	CEO	Microsoft
Bill Veghte	VP	Microsoft
Richard Stallman	founder	Free Soft..

Note:
IE is a task on its own but it can be an application of NER

Slide from Chris Brew, adapted from slide by William Cohen

Example *applications* for IE

Classified ads

Restaurant reviews

Bibliographic citations

Appointment emails

Legal opinions

Papers describing clinical medical studies

Think-Pair-Share: Tasks vs Applications

What is the difference between a task and an application?

Task: NLP community's goal of solving a single operation/problem/challenge/objective

Application: using single or multiple tasks for solving a problem; real-world use case

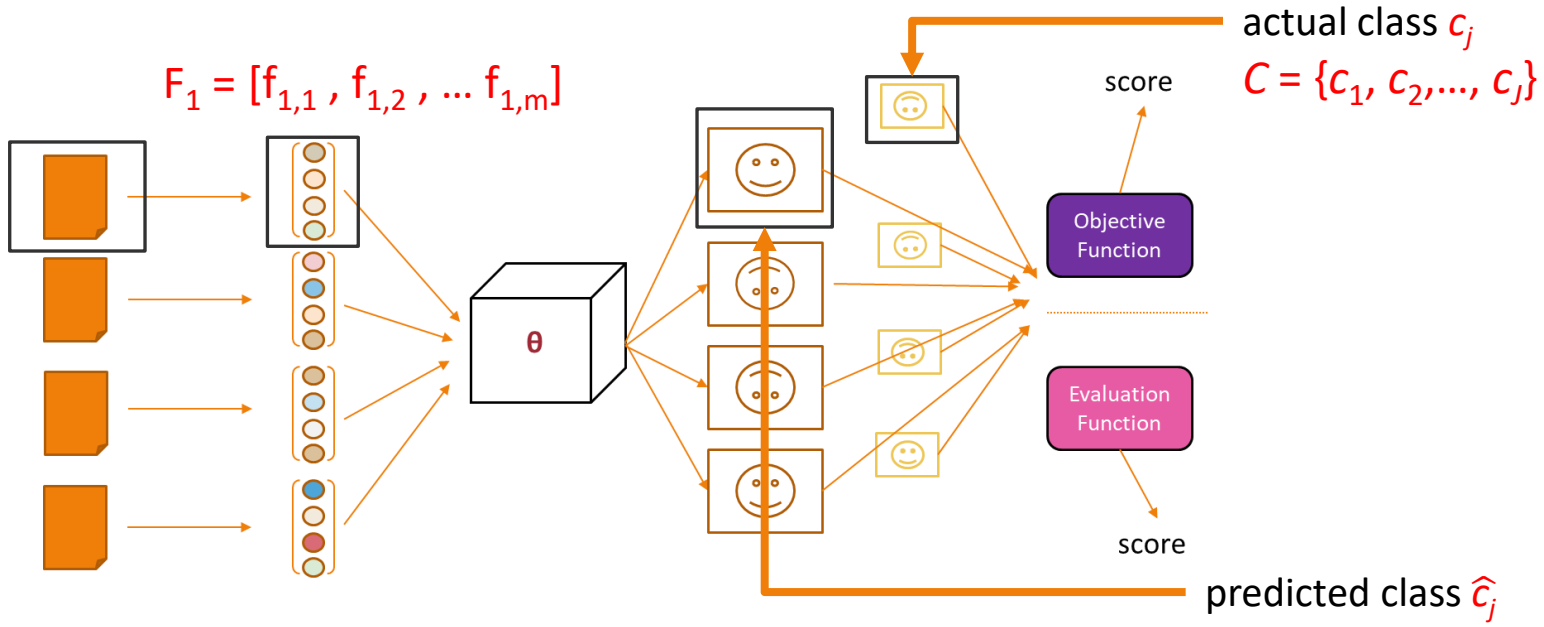
Chunking

Named entity recognition

Information extraction

Identifying idioms

features F_1 extracted from phrase p_1 and its surrounding context



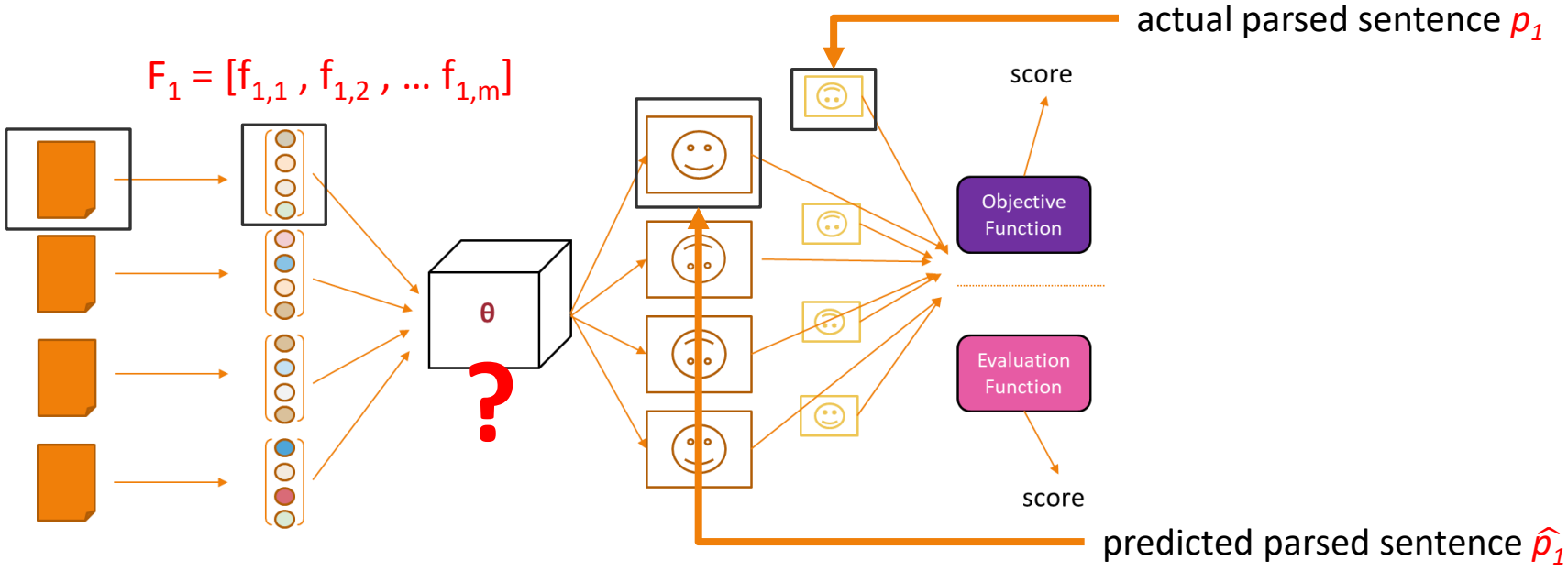
Text Annotation Tasks ("Classification" Tasks)

1. Classify the entire document ("text categorization")
2. Classify word tokens individually
3. Classify word tokens in a sequence
4. Identify phrases ("chunking")
5. Syntactic annotation (syntax parsing)
6. Semantic annotation
7. Text generation

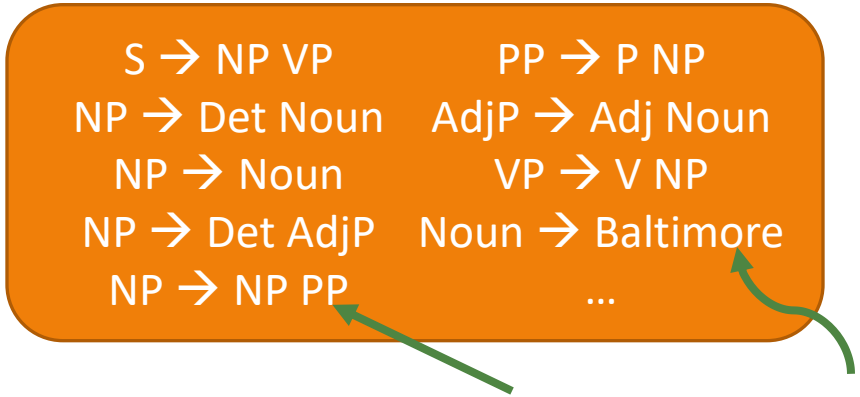
Slide courtesy Jason Eisner, with mild edits

Syntax Parsing

features F_1 extracted from phrase/sentence s_1



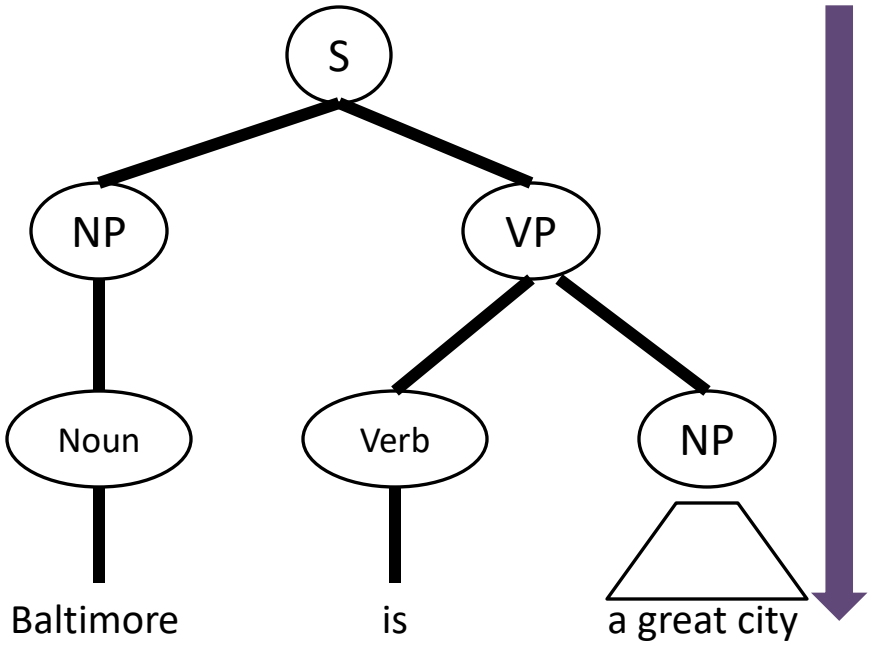
Context Free Grammar



Set of rewrite rules, comprised of non-terminals and terminals

Generate from a Context Free Grammar

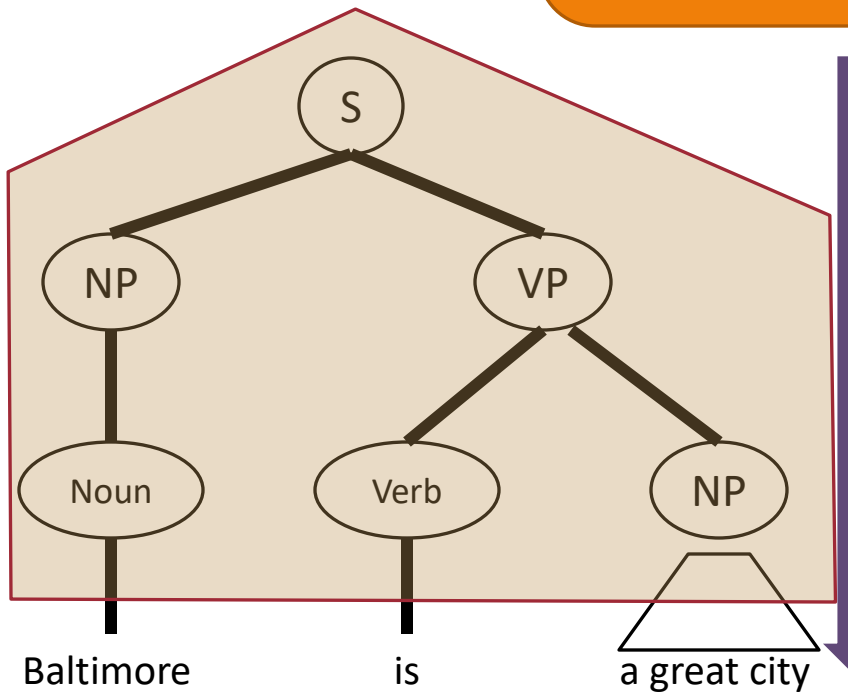
S → NP VP	PP → P NP
NP → Det Noun	AdjP → Adj Noun
NP → Noun	VP → V NP
NP → Det AdjP	Noun → Baltimore
NP → NP PP	…



Baltimore is a great city

Assign Structure (Parse) with a Context Free Grammar

S → NP VP	PP → P NP
NP → Det Noun	AdjP → Adj Noun
NP → Noun	VP → V NP
NP → Det AdjP	Noun → Baltimore
NP → NP PP	…



Baltimore is a great city

$[_S\ [_{NP}\ [_{Noun}$ Baltimore]] $[_{VP}\ [_{Verb}$ is] $[_{NP}$ a great city]]]

bracket notation

(S (NP (Noun Baltimore))
(VP (V is)
(NP a great city)))

S-expression

Why is it useful?

https://www.housebeautiful.com/uk/garden/g45582875/garden-path-ideas/

Garden Path Sentences

The old man the boat .

https://www.housebeautiful.com/uk/garden/g45582875/garden-path-ideas/

Garden Path Sentences

The old man the boat .

https://www.housebeautiful.com/uk/garden/g45582875/garden-path-ideas/

Garden Path Sentences

The rat the cat the dog chased killed ate the malt.

https://www.housebeautiful.com/uk/garden/g45582875/garden-path-ideas/

Garden Path Sentences

The rat *that* the cat the dog chased killed ate the malt.

https://www.housebeautiful.com/uk/garden/g45582875/garden-path-ideas/

Garden Path Sentences

The rat *that* the cat *that* the dog chased killed ate the malt.

https://www.housebeautiful.com/uk/garden/g45582875/garden-path-ideas/

Garden Path Sentences

The rat *that* the cat *that* the dog chased killed ate the malt.

https://www.housebeautiful.com/uk/garden/g45582875/garden-path-ideas/

Garden Path Sentences

The rat *that* the cat *that* the dog chased killed ate the malt.

https://www.housebeautiful.com/uk/garden/g45582875/garden-path-ideas/

Garden Path Sentences

The rat *that* the cat *that* the dog chased killed ate the malt.

https://www.housebeautiful.com/uk/garden/g4558287S/garden-path-ideas/

Garden Path Sentences

[The rat [the cat [the dog chased] killed] ate the malt].

Think-pair-share:
What types of applications might you use syntactic parsing for?

Language can have recursive patterns

Syntactic parsing can help identify those

Text Annotation Tasks ("Classification" Tasks)

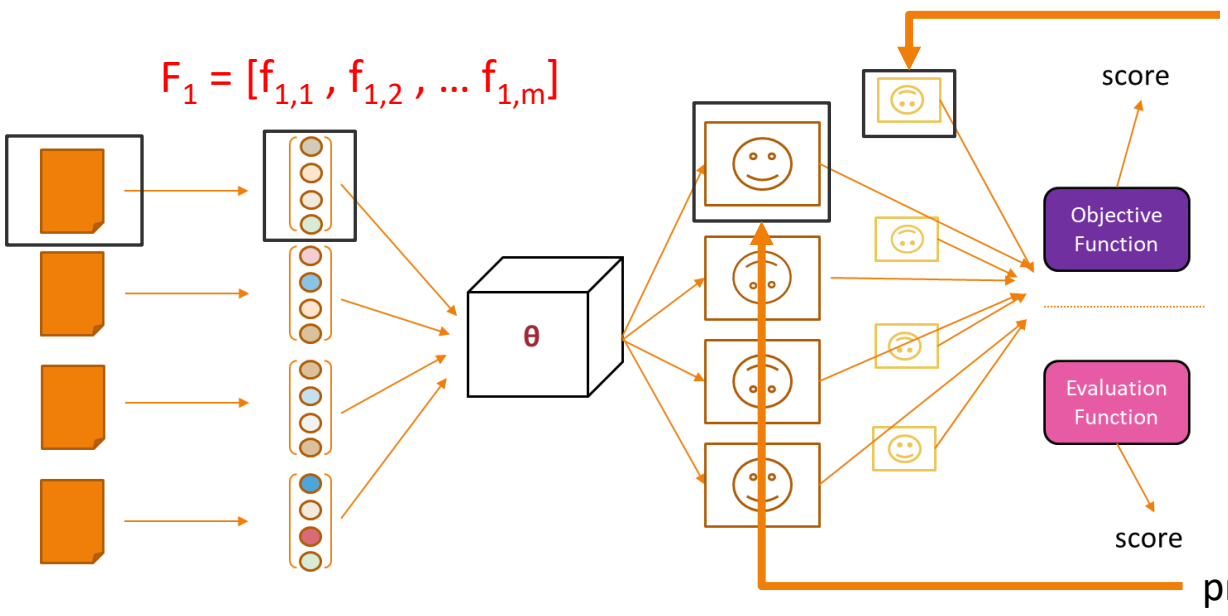
1. Classify the entire document ("text categorization")
2. Classify word tokens individually
3. Classify word tokens in a sequence
4. Identify phrases ("chunking")
5. Syntactic annotation (syntax parsing)
6. Semantic annotation
7. Text generation

Slide courtesy Jason Eisner, with mild edits

Semantic Parsing

Semantic role labeling (SRL)

features F_1 extracted from phrase/sentence s_1 and its surrounding context



actual semantic parse p_1

predicted semantic parse $\hat{p}_1$

What is a semantic parse?

Semantic Role Labeling (SRL)

For each predicate (e.g., verb)

1. find its arguments (e.g., NPs)
2. determine their **semantic roles**

John drove Mary from Austin to Dallas in his Toyota Prius.

The hammer broke the window.

- agent: Actor of an action
- patient: Entity affected by the action
- source: Origin of the affected entity
- destination: Destination of the affected entity
- instrument: Tool used in performing action.
- beneficiary: Entity for whom action is performed

Slide thanks to Ray Mooney (modified)

Semantic Role Labeling (SRL)

For each predicate (e.g., verb)

1. find its arguments (e.g., NPs)
2. determine their **semantic roles**

John drove Mary from Austin to Dallas in his Toyota Prius.

agent patient source destination instrument

What type of classification would this be?

Slide thanks to Ray Mooney (modified)

Other Current Semantic Annotation Tasks (similar to SRL)

PropBank – coarse-grained roles of verbs

NomBank – similar, but for nouns

FrameNet – fine-grained roles of any word

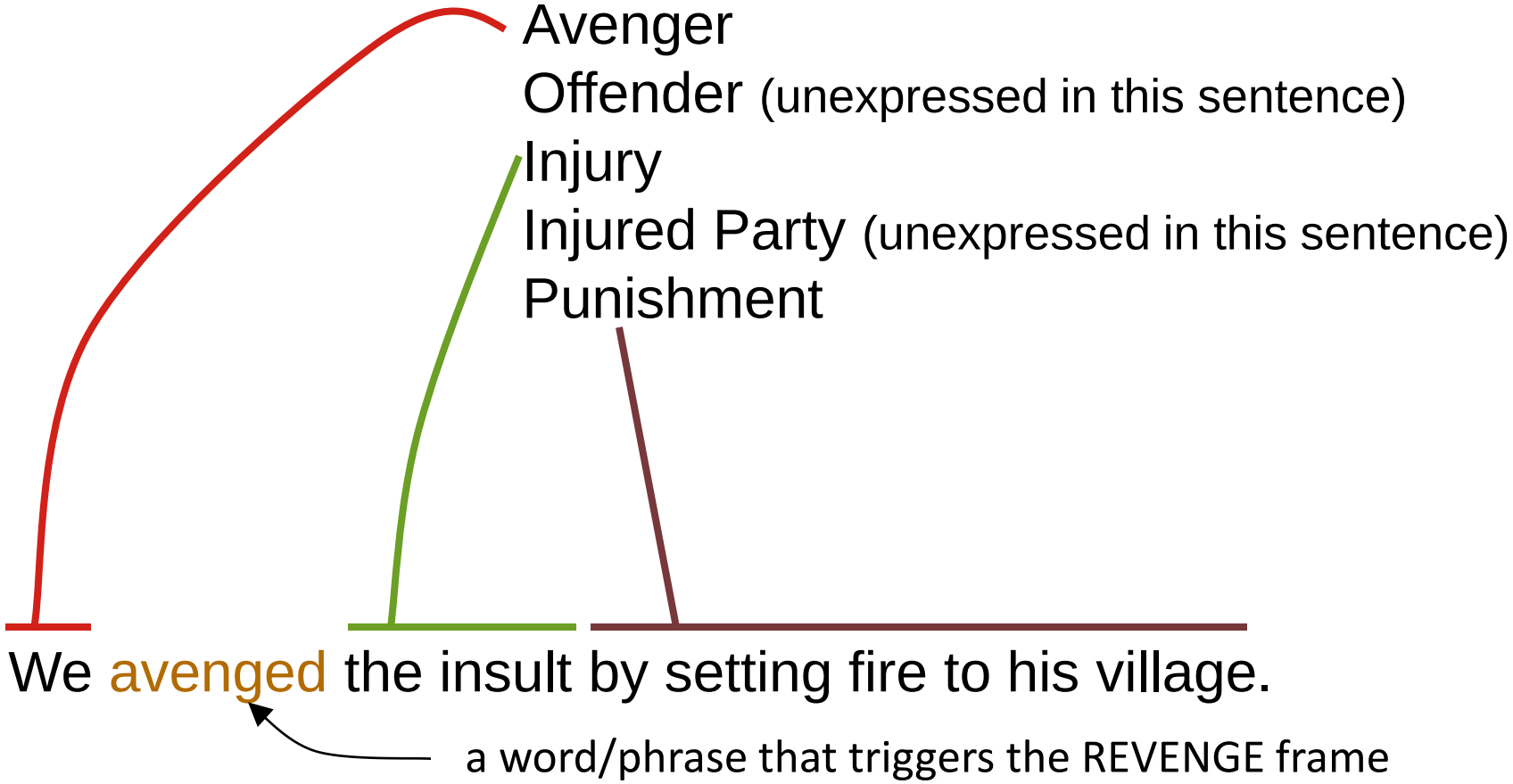
TimeBank – temporal expressions

Slide courtesy Jason Eisner, with mild edits

What type of applications might this have?

FrameNet Example

REVENGE FRAME



Slide thanks to CJ Fillmore (modified)

Text Annotation Tasks ("Classification" Tasks)

1. Classify the entire document ("text categorization")
2. Classify word tokens individually
3. Classify word tokens in a sequence
4. Identify phrases ("chunking")
5. Syntactic annotation (syntax parsing)
6. Semantic annotation
7. Text generation

Slide courtesy Jason Eisner, with mild edits

Text Generation as *Classification Problem*?

I could eat so many delicious _____

I could eat so many juicy _____

Types	Probability
apples	.03
sandwiches	.02
pineapples	.004
houses	.00002
…	…

?

Text Generation

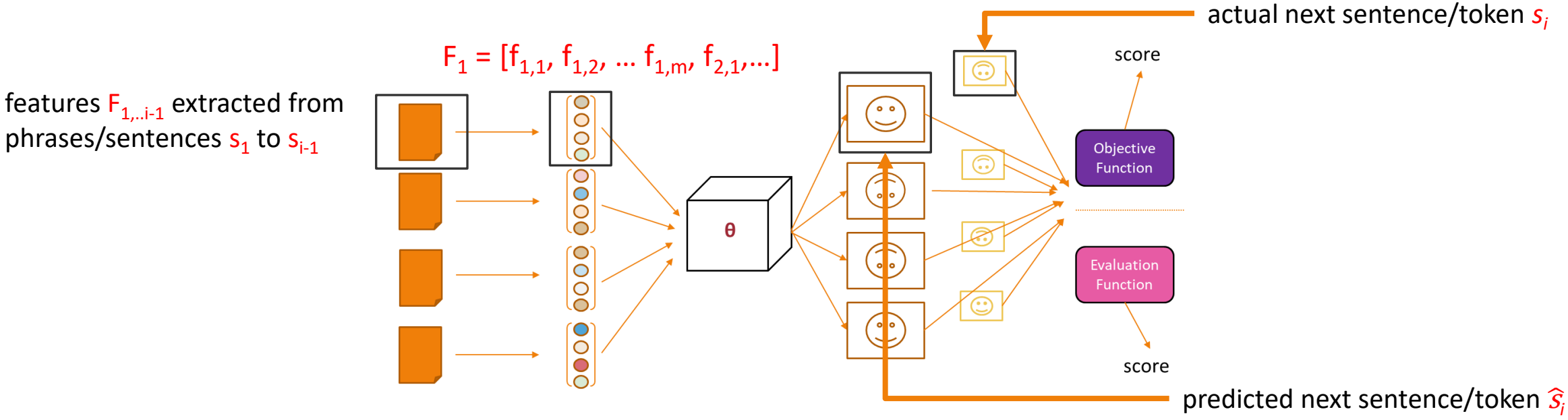
Question answering (QA)

Speech recognition (ASR)

Machine translation (MT)

Summarization

Generating text from a structured representation



There are many other tasks as well!

https://medium.com/nlplanet/two-minutes-nlp-33-important-nlp-tasks-explained-31e2caad2b1b