

# Language & IT

## Dr. Abdullah Al Fraidan

### Lecture 12- **NLP-Natural Language Processing**

#### What is NLP?

- **Natural Language Processing (NLP)**
- معالجة اللغات الطبيعية (NLP)
- Computers use (analyze, understand, generate) natural language
- يستخدم الحاسوب (التحليل والفهم، والإنشاء) للغة الطبيعية
- A somewhat applied field
- هو مجال يطبق نوعا ما
- Computational Linguistics (CL)
- اللغويات الحاسوبية
- Computational aspects of the human language faculty
- الجوانب الحسابية في كلية اللغة البشرية
- More theoretical
- المزيد من النظرية

#### Why Study NLP?

- Human language interesting & challenging
- اللغة البشرية مثيرة للاهتمام والتحدي
- NLP offers insights into language
- NLP يقدم نظرة ثاقبة للغة
- Language is the medium of the web
- اللغة هي وسيلة من الويب
- Interdisciplinary: Ling, CS, psych, math
- التخصصات: لينغ، CS، النفس، الرياضيات

- Help in communication

• يساعد في الاتصالات

- With computers (ASR, TTS)

• مع الكمبيوتر (ASR، TTS)

- With other humans (MT)

• مع غيره من البشر (MT)

- Ambitious yet practical

• الآن الطموح عملي

### - الأهداف - Goals of NLP

#### • Scientific Goal

• الهدف العلمي

- Identify the computational machinery needed for an agent to exhibit various forms of linguistic behavior

• تحديد الآلية الحسابية اللازمة كعامل لعرض أشكال مختلفة من السلوك اللغوي

#### • Engineering Goal

• الهدف الهندسي

- Design, implement, and test systems that process natural languages for practical applications

• تصميم وتنفيذ واختبار النظم لمعالجة اللغات الطبيعية للتطبيقات العملية

### - التطبيقات - Applications

#### speech processing:

get flight information or  
book a hotel over the phone

معالجة الكلام: الحصول على  
معلومات الطيران أو حجز فندق عبر  
الهاتف

#### information extraction:

discover names of people  
and events they participate  
in, from a document

استخراج المعلومات: اكتشاف أسماء  
الأشخاص والأحداث المشاركة، من  
الوثيقة

#### machine translation:

translate a document  
from one human  
language into another

الترجمة الآلية: ترجمة مستند من  
لغة بشرية واحدة إلى أخرى

question answering: find answers to  
natural language questions in a text  
collection or database

ردا على السؤال: العثور على إجابات لأسئلة اللغة  
الطبيعية في مجموعة نص أو قاعدة بيانات

summarization: generate a short  
biography of Noam Chomsky from one or  
more news articles

التلخيص: إنشاء سيرة ذاتية قصيرة من نعوم تشومسكي  
من أحد أو أكثر المقالات الإخبارية

## - المواضيع العامة - General Themes

- Ambiguity of Language
  - Language as a formal system
  - Rule-based vs. Statistical Methods
  - The need for efficiency
- غموض اللغة
  - اللغة كنظام رسمي
  - أساليب قائمه على القواعد مقابل الإحصائية
  - الحاجة إلى تحقيق الكفاءة

## غموض اللغة Ambiguity of language

- Phonetic
  - [ralt] = write, right, rite
  - Lexical
  - can = noun, verb, modal
  - Structural
  - I saw the man with the telescope
  - Semantic
  - dish = physical plate, menu item
  - All of these make NLP difficult
- الصوتي
  - المعجمي
  - الهيكلي
  - الدلالي
  - كل هذه تجعل من الصعب NLP

## - اللغة كنظام رسمي - Language as a formal system

- We can treat parts of language formally
  - Language = a set of acceptable strings
  - Define a model to recognize/generate language
- إننا يمكن التعامل أنحاء اللغة رسميا
  - اللغة = مجموعة من الاوتار المقبولة
  - تعريف نموذج للاعتراف / ينشئ اللغة

- Works for different levels of language (phonology, morphology, etc.)
- يعمل من أجل مستويات مختلفة من اللغة (علم الأصوات، علم الصرف، الخ.)
- Can use finite-state automata, context-free grammars, etc. to represent language
- يمكن استخدام الآلي ومحدد للدولة، قواعد خالية من السياق، وما إلى ذلك لتمثيل اللغة

### Rule-based & Statistical Methods

#### أساليب التقيد بالقواعد والإحصائية

- Theoretical linguistics captures abstract properties of language
- تلتقط اللغويات النظرية خصائص مجردة من اللغة
- NLP can more or less follow theoretical insights
- NLP يمكن أكثر أو أقل لأتباع أفكار النظرية
- Rule-based: model system with linguistic rules
- المستندة إلى القاعدة: النظام النموذجي مع القواعد اللغوية
- Statistical: model system with probabilities of what normally happens
- الإحصائية: نظام نموذجي مع احتمالات ما يحدث عادة
- Hybrid models combine the two
- نماذج الهجينة للجمع بين الاثنين

### -الحاجة إلى تحقيق الكفاءة The need for efficiency

- Simply writing down linguistic insights isn't sufficient to have a working system
- ببساطة تدوين أفكار لغوية غير كافيا لأن يكون هناك نظام عمل
- Programs need to run in real-time, i.e., be efficient
- تحتاج البرامج للتشغيل في الوقت الحقيقي، أي أن تكون فعالة
- There are thousands of grammar rules which might be applied to a sentence
- هناك الآلاف من القواعد النحوية التي يمكن تطبيقها للجملة
- Use insights from computer science
- استخدام الأفكار من علوم الكمبيوتر
- To find the best parse, use chart parsing, a form of dynamic programming
- للعثور على أفضل تحليل، واستخدام تحليل الرسم البياني، شكلا من أشكال البرمجة الديناميكية