

بسم الله الرحمن الرحيم

# پردازش زیانهای طبیعی

جلسه ۱

احسان الدین عسگری

مهر ۱۴۰۲

<http://language.ml/>

دانشکده مهندسی کامپیوتر

آزمایشگاه پردازش هوشمند متن و زبان و علوم انسانی محاسباتی

asgari@berkeley.edu



- زبان



# تعريف زبان

• معرفی

• زبان

• پردازش زبان طبیعی

• درس NLP

• جلسه بعد

Chomsky (1959: 137) “*A language is a collection of sentences of finite length all constructed from a finite alphabet (or, where our concern is limited to syntax, a finite vocabulary) of symbols.*”



# تعریف زبان

• معرفی

• زبان

• پردازش زبان طبیعی

• NLP درس

• جلسه بعد

Chomsky (1959: 137) “*A language is a collection of sentences of finite length all constructed from a finite alphabet (or, where our concern is limited to syntax, a finite vocabulary) of symbols.*”



## DNA Language

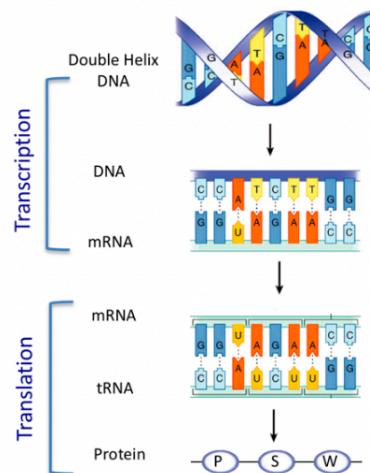
Sentences out of {A,T,C,G}

## RNA Language

Sentences out of {A,U,C,G}

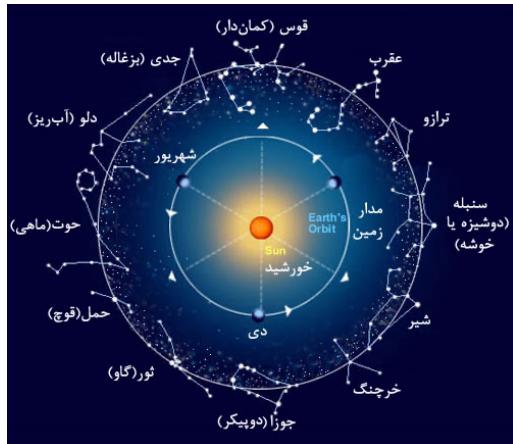
## Protein Language

Sentences out of  
{A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,V,W,X,Y}



# تعریف زبان

Chomsky (1959: 137) “*A language is a collection of sentences of finite length all constructed from a finite alphabet (or, where our concern is limited to syntax, a finite vocabulary) of symbols.*”

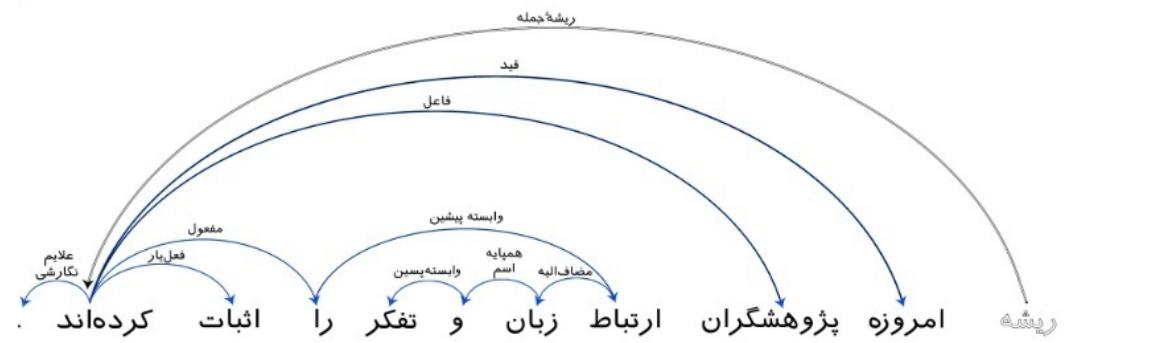


- معرفی زبان
- پردازش زبان طبیعی NLP درس
- جلسه بعد

# تعریف زبان



- مثال ارتباطات درختی و شبکه‌ای



- معرفی زبان
- پردازش زبان طبیعی
- درس NLP
- جلسه بعد

# تعریف زبان

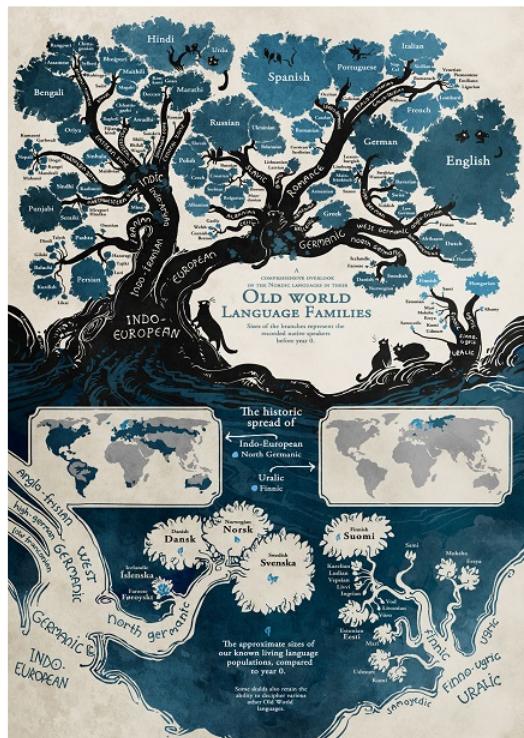
Chomsky (1959: 137) “*A language is a collection of sentences of finite length all constructed from a finite alphabet (or, where our concern is limited to syntax, a finite vocabulary) of symbols.*”



مجموعه توصیفاتی از حقایق: توصیفاتی در محور زمان از المان‌های تکرار پذیر که با هم ارتباطات تنگاتنگ ساختاری و معنایی دارند: ارتباطات درختی و شبکه‌ای.

- معرفی
- زبان
- پردازش زبان طبیعی
- NLP
- درس
- جلسه بعد

# پردازش زبان انسان



◦ حدود ۷۰۰۰ زبان طبیعی

<https://wals.info>

◦ بسیاری با محدودیت منابع

◦ چرا به تکنولوژی زبانی نیاز داریم؟

- معرفی زبان
- پردازش زبان طبیعی
- درس NLP
- جلسه بعد

# زبانهای ایرانی

سامانه جامع معرفی زبانهای ایرانی و منابع زبانی و زبانشناسی  
آزمایشگاه پردازش هوشمند علت و زبان و علوم انسانی محاسباتی



# - پردازش زبان طبیعی



• معرفی

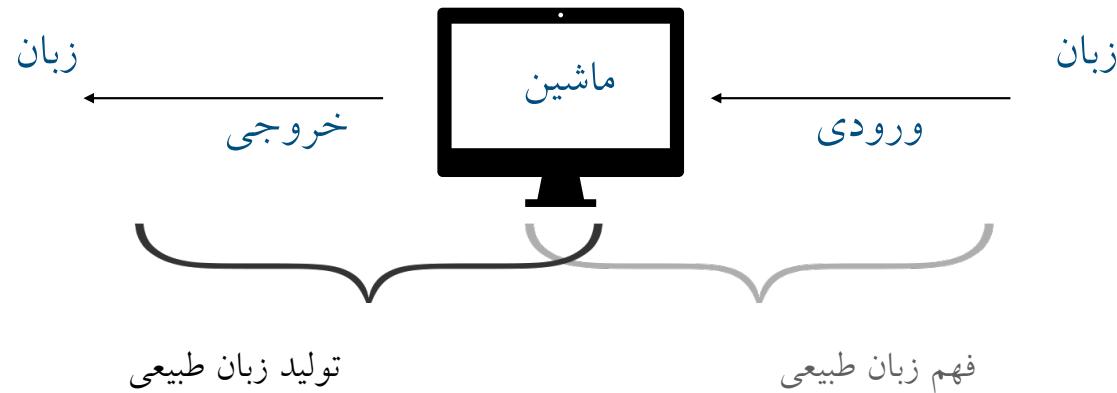
• زبان

• پردازش زبان طبیعی

• درس NLP

• جلسه بعد

# پردازش زبان طبیعی (NLP)



Natural Language generation

Natural Language Understanding

چگونه فهمی؟ پردازشی که به نمایش قابل فهم برای ماشین بیانجامد.

# پردازش زبان طبیعی (NLP)

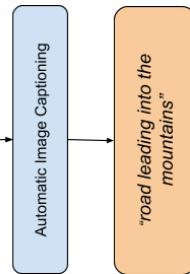
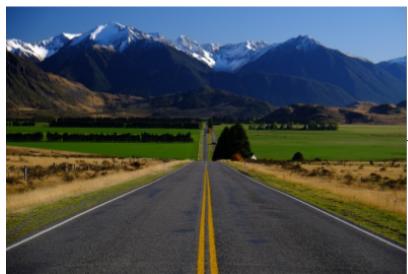
- مطالعه علمی زبان با دید رایانشی
- کار با مدل‌های محاسباتی پدیده‌های مختلف زبانی با انگیزه علمی یا فنی.

[www.aclweb.org](http://www.aclweb.org)

- فرق میان NLP و Computational Linguistics

- معرفی زبان
- پردازش زبان طبیعی
- درس NLP
- جلسه بعد

# کاربردهای NLP



**Contextual Spelling Error**

Deer Mr. Theodore: **Spelling Error**

I am exceedingly interested in this po  
and employment background are app

While working toward my degree, I wa  
small firm. I increased my call volume  
success. I will completes my degree in  
employment in early June.

**Grammar Error**

A screenshot of a mobile application interface. At the top, it says "Contextual Spelling Error" and "Deer Mr. Theodore: Spelling Error". Below that, there is a red underline over the word "exceedingly" in the sentence "I am exceedingly interested in this po...". Further down, another red underline is over the word "completes" in the sentence "I will completes my degree in...". At the bottom, it says "Grammar Error". The overall context is a job application or resume.

- معرفی
- زبان
- پردازش زبان طبیعی
- درس NLP
- جلسه بعد

# NLP سایر مسائل کاربردی

- Spelling Suggestions
- Grammar Checking
- Synonym Generation
- Information Extraction
- Text Categorization
- Summarization
- Essay Scoring

- Automated Customer Service
- Text generation
- Machine Translation
- Question Answering
- Improving Web Search Engine results
- Automated Metadata Assignment
- Online Dialogs

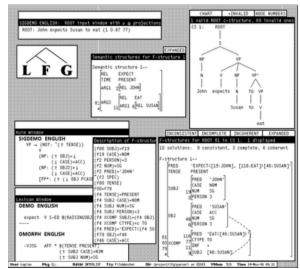
• معرفی زبان پردازش زبان طبیعی

• درس NLP جلسه بعد

- معرفی زبان پردازش زبان طبیعی
- درس NLP
- جلسه بعد

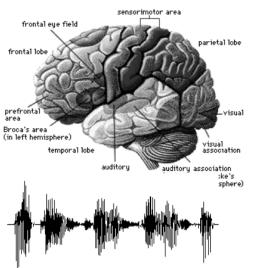
# پردازش زبان طبیعی - زمینه میان رشته ای

- علوم کامپیوتر
- یادگیری ماشین: مدلها آماری یا شبکه های عصبی
- زبان شناسی
- علوم انسانی و اجتماعی



## رشته های هم مرز

- زبان شناسی
- پردازش گفتار (صوت)
- علوم شناختی
- یادگیری ماشین



## هدف گیری درس

- آشنایی با مسائل مهم پردازش زبان طبیعی
- تجربه عملی کار با داده‌های متنی و پیش‌پردازشها
- تئوری روش‌های اصلی SOTA
- پرداختن به پردازش زبان‌های ایرانی و مشارکت در توسعه بنچمارک
- در پروژه‌ها) پرداختن به مسائل نو و کاربردی - قابل چاپ



## ماژولهای درس



مسائل کاربردی دیگر



تولید متن

۴ جلسه



طبقه‌بندی کلمه  
و پرسش و پاسخ

۴ جلسه



طبقه‌بندی متن

۴ جلسه



مدلهای زبانی

۵ جلسه



آشنایی با متن و کلمات  
روشهای پیش‌پردازش

۷ جلسه

۲ تمرین

پایان‌ترم

۱ تمرین

۱ تمرین

۱ تمرین

۱ تمرین

۱ کوییز

۱ کوییز

۱ کوییز

۱ کوییز

۱ کوییز



# ارزیابی

- ۱۲۰ نمره (برای ۶ تمرین)
- ۴۰ نمره پروژه پایانی
- ۲۰ نمره کوییزها (از هر مژول)
- ۲۰ نمره پایان‌ترم
- تا سقف ۱ نمره حضور فعال و مستمر در کلاس
- تا سقف ۵.۰ نمره ارائه اختیاری صحنه‌ها

– البته توجه داشته باشید که ممکن است نمره‌های برخی بخش‌ها در انتهای ترم در نرمال‌سازی تغییراتی داشته باشند.

- معرفی
- زبان
- پردازش زبان طبیعی
- درس **NLP**
- جلسه بعد