Why NLP?

Natural Language Processing
CS 4120/6120—Spring 2017
Northeastern University

David Smith
Codes
Completely Automated Public Turing test to tell Computers and Humans Apart
Fine, walk away. I'm gonna go cry into a pint of Ben&Jerry's Brownie Batter(tm) ice cream [link], then take out my frustration on a variety of great flash games from PopCap Games(r) [link].
One thing I wanted to ask you about is this. A most serious problem, for UNESCO and for the constructive and peaceful future of the planet, is the problem of translation, as it unavoidably affects the communication between peoples. Huxley has recently told me that they are appalled by the magnitude and the importance of the translation job.

Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worth while.

Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography—methods which I believe succeed even when one does not know what language has been coded—one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: “This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.”
ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO THE ENTSCHEIDUNGS PROBLEM

By A. M. Turing.

[Received 28 May, 1936.—Read 12 November, 1936.]

The “computable” numbers may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means. Although the subject of this paper is ostensibly the computable numbers, it is almost equally easy to define and investigate computable functions of an integral variable or a real or computable variable, computable predicates, and so forth. The fundamental problems involved are, however, the same in each case, and I have chosen the computable numbers for explicit treatment as involving the least cumbrous technique. I hope shortly to give an account of the relations of the computable numbers, functions, and so forth to one another. This will include a development of the theory of functions of a real variable expressed in terms of computable numbers. According to my definition, a number is computable if its decimal can be written down by a machine.

In §§ 9, 10 I give some arguments with the intention of showing that the
ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO
THE ENTSCHEIDUNGSPROBLEM

By A. M. Turing.

[Received 28 May, 1936.—Read 12 November, 1936.]

The "computable" numbers may be described briefly as the real
numbers whose expressions as a decimal are calculable by finite means.
Although the subject of this paper is ostensibly the computable numbers,
it is almost equally easy to define and investigate computable functions
of an integral variable or a real or computable variable, computable
predicates, and so forth. The fundamental problems involved are,
however, the same in each case, and I have chosen the computable numbers
for explicit treatment as involving the least cumbersome technique. I hope
shortly to give an account of the relations of the computable numbers,
functions, and so forth to one another. This will include a development
of the theory of functions of a real variable expressed in terms of com-
putable numbers. According to my definition, a number is computable
if its decimal can be written down by a machine.

In §§ 9, 10 I give some arguments with the intention of showing that the

with the m-configuration written below the scanned symbol. The
successive complete configurations are separated by colons.

```
: o o 0 : o o 0 0 : o o 0 0 : o o 0 0 : o o 0 1:
 o o 0 1 : o o 0 0 1 : o o 0 0 1 : o o 0 0 1:
 o o 0 1 : o o 0 1 : o o 0 0 1 0:
 o o 0 1 0 : ...:
```

This table could also be written in the form

```
: o o 0 0 : o o 0 0 : .... 
```

(C)
TENSE AND MOOD IN INDO-EUROPEAN SYNTAX*

1. The historical present

The 'historical' or 'dramatic' present tense used in narrating past events, which is common in many Indo-European languages, has always been interpreted in essentially semantic terms. A typical traditional formulation is it is quite mistaken to transfer it to the earlier stages of Indo-European. In Greek, Old Irish, and Old Norse, for example, the historical present has quite different syntactic and semantic properties, to which the traditional idea, or any of its variants², must utterly fail to do justice.

* This work was supported in part by the Joint Services Electronics Program under Contract DA36-039-AMC-03200(E); in part by the National Science Foundation (Grant GP-2495), the National Institutes of Health (Grant MH-04737-05), the National Aeronautics and Space Administration (Grant NsG-496), and the U.S. Air Force (ESD Contract AF 19 (628)-2487). – I thank Michael Connolly, Eric Hamp, Einar Haugen, George Lakoff, Calvert Watkins, and Roy Wright for offering valuable criticism and/or referring me to some of the examples cited here.
The Turing Test
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?
Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?

Witness: In a way.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn’t scan.

Interrogator: How about "a winter's day," That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?

Witness: In a way.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?

Witness: In a way.

Interrogator: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?

Witness: In a way.

Interrogator: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.
The Turing Test

Interrogator: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day," That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?

Witness: In a way.

Interrogator: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.

Witness: I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.
Noam Chomsky
Noam Chomsky

- Anarcho-syndicalist polemicist
Noam Chomsky

• Anarcho-syndicalist polemicist

• Inventor of several theories of “generative grammar”
Noam Chomsky

• Anarcho-syndicalist polemicist

• Inventor of several theories of “generative grammar”

• Pioneer of formal language theory
Modularity
Linguistic Modules

- Phonetics and phonology
- Morphology
- Syntax
- Semantics
- Pragmatics
- Discourse

With lots of crossings between levels!
Phonetics and Phonology

• Phonetics: language sounds & their physiology

• Phonology: systems of discrete sounds in languages

• E.g.: devoicing of *it is* to *it’s*

• E.g.: syllable structure: *sign, signify*
Morphology

- Inflectional (in some languages):
  - love → loved

- Derivational:
  - tea-cup, un-helpful, with-stand, craisin

- Turkish: uygarlastiramadiklarimizdanmissinizcasina
  - uygar las tir ama dik lar imiz dan mis siniz casina

- (behaving) as if you are among those whom we could not civilize
Morphological Tagging

There are many kinds of trench mortars.

c. Klimatizovaná jídelna, světlá místnost pro snídaně.

Air-conditioned dining room,
Syntax
Syntax

Constituency
Syntax

IP

NP

Speaking

VP

V

changes

NP

language

Constituency

Dependency

SUBJ

Verbing

weirds

OBJ

language
Semantics

Pierre Vinken, 61 years old, will join the board as a nonexecutive director.
Semantics

Pierre Vinken, 61 years old, will join the board as a nonexecutive director.

PropBank join predicate

<table>
<thead>
<tr>
<th>ARG0</th>
<th>ARG1</th>
<th>ARG-PRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vinken</td>
<td>board</td>
<td>director</td>
</tr>
</tbody>
</table>
Pragmatics

- Context affects meaning
- Conversational implicature
  - *May I speak to your mother? Yes.*
- Speech acts: “how to do things with words”
  - *I grant you permission to speak.*
Discourse

- Study of units larger than a single utterance
  - Turn taking
- Coreference
- Organized exposition
It All Hangs Together

Phonology

Inflectional Morphology

Morphology

Syntax

Pragmatics

Syntax

Discourse

Semantics

Le langage est un système où tout se tient. — Meillet
Applications
<bookstore>
  <book category="COOKING">
    <title lang="en">Everyday Italian</title>
    <author>Giada De Laurentiis</author>
    <year>2005</year>
    <price>30.00</price>
  </book>
  <book category="CHILDREN">
    <title lang="en">Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>
  <book category="WEB">
    <title lang="en">Learning XML</title>
    <author>Erik T. Ray</author>
    <year>2003</year>
    <price>39.95</price>
  </book>
</bookstore>
<bookstore>
  <book category="COOKING">
    <title lang="en">Everyday Italian</title>
    <author>Giada De Laurentiis</author>
    <year>2005</year>
    <price>30.00</price>
  </book>
  <book category="COOKING">
    <title lang="en">Harry Potter and the Half-Blood Prince</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>
  <book category="WORLD">
    <title lang="en">Ratcatcher</title>
    <author>Erik Tulving</author>
    <year>2003</year>
    <price>39.95</price>
  </book>
</bookstore>
2. Cunctae aequales. 16. Et hic quidem punctus, centrum circuli dicitur.
3. Diametrum circuli, est linea recta, que super eius centrum transiens,
4. extremitateque suae circuteretig applicata, circuli in duos media divi-
5. ditur. 18. Semicirculus, est figura plana diametro circuli, & mediate
circuteretig constans. 20. Portio circuli, est figura plana, recta linea &
6. parte circumferentiae constas, semicirculo quod aut maior aut minor.
8. Semicirculus.
10. Rectilinear figure sunt, que rectis lineis continentur. 21. Quadrum
11. que dexter tria latera, que tribus rectis lineis: 22. Quudem quadrilatero,
12. quatuor rectis lineis: 23. Quaqua multilatera, que pluralibus quatuor
13. rectis lineis continet. 24. Figurarum tria latera, alia est tria latera
14. aequitatis: 25. Alia tria latera, duobus, et quatuor lineis aequitatis
15. tria latera aequitatis: 26. Alia tria latera, tria, et quatuor lineis aequitatis
16. tria latera, et quatuor lineis aequitatis. 27. Habes iterum alia est orthogoniuni
17. cum sinus recti anguli habes. 28. Alia est oblonga ungu, que quatuor
18. anguli aequitant. 29. Alia est oblonga ungu, que tres anguli sunt acuti.
20. Figurarum autem quadrilateralium, alia est quadratum, quod est aequi-
21. laterum circulii: 31. Alia est tetrugonum longum, que est figura recta
22. aequilateralis, sed qua quaedam recta aequilateralis: 32. Alia est helmaeun, que est equilatera,
23. semicirculo autem, que opposita late
25. circulus aequabilis aequilibus aequitatis, ide
26. tamen nec rectis angulis nec aequi laterib, continetur.
27. Parte.
GEOMET. ELEMENT. EVCLIDIS
functæguales. 16. Et hic quidem punctus, centrum circuli dicitur.
21. Diameter circuli, est linea recta, que super eius centrum transitens,
extrimateque sua circulture applicatis, circuli in duo media diuidit.
20. Semi-circuler, est figura plana diametris circuli, &e mediate
circulture constet. 21. Portio circuli, est figura plana, recta linea &
parte circularentur extrema, semicirculer quod aut maior aut minor.


Semirculer.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Minor portio.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.

Rectilineæ figura sunt, quæ rectis lineis continnuntur. 21. Quamur
quadrilatera, que tribus rectis lineis: 22. Quaedam quadrilatera,

Quadrula.
<table>
<thead>
<tr>
<th>Name</th>
<th>Message/Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roland R. Duncan</td>
<td>Show your love to your special people! EMAIL ID: ReJr</td>
</tr>
<tr>
<td>Doctor Max Stokes</td>
<td>[ JHSPAM-ALERT-IP ] It's important for you - XMAILOE %XMIMEOE Make your lady-love contented! You</td>
</tr>
<tr>
<td>Cary Phelps</td>
<td>No Hassle Business Loans - If you have your own business and need IMMEDIATE money to spend ANY way y</td>
</tr>
<tr>
<td>Elias Moran</td>
<td>[ JHSPAM-ALERT-IP ] Loans - Small Business - If you have your own business and require IMMEDIATE ready money to spend ANY way y</td>
</tr>
<tr>
<td>Dr John Avery</td>
<td>[ JHSPAM-ALERT-IP ] Something interesting for you - XMAILOE %XMIMEOE X-Antivirus: avast! (VPS 08</td>
</tr>
<tr>
<td>Dr John Avery</td>
<td>[ JHSPAM-ALERT-IP ] Something interesting for you - XMAILOE %XMIMEOE X-Antivirus: avast! (VPS 08</td>
</tr>
<tr>
<td>Roland Reid</td>
<td>Business Loans - If you have your own business and need IMMEDIATE ready money to spend ANY way y</td>
</tr>
<tr>
<td>peggy.ruby</td>
<td>[ JHSPAM-ALERT-IP ]</td>
</tr>
<tr>
<td>Dr John Avery</td>
<td>[ JHSPAM-ALERT-IP ] It's important for you - XMAILOE %XMIMEOE Make your lady-love contented! You</td>
</tr>
<tr>
<td>Thermometer</td>
<td>°C cosystem &amp; THERMOMETER - 30% OFF -療具、康體 一切彌足珍貴的健康設備 Access - 50% OFF -療具、康體 一切彌足珍貴的健康設備</td>
</tr>
<tr>
<td>[email]</td>
<td>ếp &amp; ツ Overview of the site: - <a href="http://www.thermometer.com">http://www.thermometer.com</a></td>
</tr>
</tbody>
</table>
Would you like to...

Add to calendar
Feature space Maximu...
Fri Feb 8 12pm – Fri F...
الادعاء يطلب أقصى عقوبة لثلاثة متهمين بالتخطيط لاغتيال بوش

عمان الحياة - 08/02/07

طالب مدعى عام محكمة أمن الدولة الأردنية بتوفيق أقصى عقوبة ممكنة على ثلاثة متهمين بالتخطيط لاغتيال الرئيس الأمريكي جورج بوش أثناء زيارته عمان في تشرين الثاني (نوفمبر) 2006. بعدما وجه إليهم اتهامات ب beğنها «القيام بأعمال إرهابية، واستخدام مواد مهلكة، وحمل وحيازة أسلحة أتوماتيكية ومفرقعات» من دون ترخيص بقصد استعمالها على وجه غير مشروع.

وتل يرئ المحكمة تقريراً طبياً صادراً عن لجنة قيادة الوضع العلوي للمتهم_statuses_الزاهرة الذي طلب محاميته إجلاءه على لجنة لمعرفة مدى إدراكه لطبيعة أفعاله وإمكان ممارسته أمام المحكمة. وآخذت التقرير أن المتهم مدرك لأفعاله وأقواله. وطلب وكيل الدفاع المحامي عبد الكريم الشريرة من المحكمة إجلاءه لتقريب زواجه في Nhiềubing 12 من الشهر المقبل، في القوات الدقيقة، والتخصيص لثلاثة أيام، 08.
 Prosecution requests the maximum penalty for the three accused of planning to assassinate Bush

Oman life - 07/02/08

Asked prosecutor in the state security court of Jordan signed the maximum possible sentence to three accused of planning to assassinate President George Bush during his visit to Amman in November (November) 2006, after it drew them was charged with «terrorist acts, the use of flammable materials, and carrying and possessing automatic weapons and explosives without a licence in order to use illegal». The President of the Court read out a medical report issued by the Commission assessed the mental status of the accused Azwahrv Stam, who asked his lawyer referred to a committee to determine the extent aware of the nature of his acts and the possibility of his appearance before the court. The report proved that «the accused was aware of his actions and sayings». Asked assistant defense lawyer court delayed by the court to submit a report on the matter of the accusations against them.
Translation
Er wird in den Strassen wandern
Translation

Er wird in den Strassen wandern

He will in the streets walk
Translation

Er wird in den Strassen wandern

He will in the streets walk

He will walk in the streets
Er wird in den Strassen wandern

He will in the streets walk

He will walk in the streets

Er wird in den kleinen Strassen wandern
Translation

Er wird in den Strassen wandern

*He will in the streets walk*

Er wird in den *kleinen* Strassen wandern

*He will in the *small* streets walk*
Translation

Er wird in den Strassen wandern
*He will in the streets walk*

He will walk in the streets

Er wird in den *kleinen* Strassen wandern
*He will in the small streets walk*

He is in the small streets hike
Translation

Er wird in den Strassen wandern
He will in the streets walk
He will walk in the streets

Er wird in den kleinen Strassen wandern
He will in the small streets walk
He is in the small streets hike
Question Answering

Who is the leader of France?
Who is the leader of France?

Henri Hadjenberb, who is the leader of France’s Jewish community.
Who is the leader of France?

Henri Hadjenberb, who is the leader of France’s Jewish community.

Bush met with French president Jacques Chirac.
Multilingual “Topics”
European Parliament Corpus
Multilingual “Topics” Wikipedia comparable articles
world ski km won
world ski km won

actor role television actress
world ski km won
actor role television actress
ottoman empire khan byzantine
Projecting Hidden Structure

Annotations From Existing English Tools

Induced Annotations for Chinese
NLP Tasks

• Analog to digital
  • OCR, Speech Recognition
• Individual language modules
  • Morphology, Syntax, Semantics, and Discourse
• Language to data
  • Information extraction and retrieval
• Language to language
  • Translation, summarization, dialogue systems
Monolingual & Multilingual

- Analysis technologies for languages
  - Morphology, syntax, semantics
- Translation technologies
  - Dictionaries, cross-lingual IR, MT
- Multilingual exploratory data analysis
  - Clustering, classification→model building
A Few Problems
Morphological Ambiguity

There are many kinds of trench mortars.

c. Klimatizovaná jídelna, světlá místnost pro snídaně.

Air-conditioned dining room,
Syntactic Ambiguity
More Ambiguity

- Iraqi Head Seeks Arms
- Juvenile Court to Try Shooting Defendant
- Teacher Strikes Idle Kids
- Stolen Painting Found by Tree
- Kids Make Nutritious Snacks
- Local HS Dropouts Cut in Half
- British Left Waffles on Falkland Islands
- Red Tape Holds Up New Bridges
- Clinton Wins on Budget, but More Lies Ahead
- Ban on Nude Dancing on Governor’s Desk
Why is NLP Hard?

• The rules are ambiguous
• We don’t know the rules
• We need to combine lots of weak evidence
• It’s AI complete
• Language is nearly co-extensive with humanity
• To the rescue: probability, machine learning
Why is NLP in CS?

• How about...

• Linguistics

• Statistics

• Psychology and Cognitive Science

• The Lang/Lit Humanities

• All of the above!

• Focus on algorithms, data analysis, engineering
What You’ll Learn in NLP

• Looking at data
  • Phenomena and problems
• Modeling data
  • Linguistic and statistical tools
• Algorithms and implementation
  • Efficient computation, practical systems
No Really, What’ll I Learn?

- Models of language
  - n-grams, grammars, generative, discriminative
- Algorithms to tame complexity
  - Finite-state models, regular expressions
  - Context-free grammars and parsers
  - Backpropagation and learning
- Problem solving: classification, structured prediction, generation
Who – Where – When

• Instructor: David Smith
  • dasmith@ccs.neu.edu
  • WVH 356, TBA or by appointment

• TA: Liwen Hou
  • lhou@ccs.neu.edu
  • WVH 472, TBA

• Thursdays 6-9, Behrakis 325
  • www.ccs.neu.edu/course/cs6120sp16
What

• Upper undergrad & grad course in NLP
  • Learning to *read papers* in NLP
• Presentation and participation (20%)
• Homework assignments (3/4 for 30%)
• Final project (50%)
  • Application of a model to different data
  • Report analyzing implementation, results
More on Grading

• All code and written words must be your own

• **OK**: Discussing problems with others, as long as you acknowledge it

• **Not OK**: Copying other text or homework, divide-and-conquer coding
What

- Lectures introduce algorithms, models, learning methods
- Background reading in two books:
  - *Speech and Language Processing*. Jurafsky & Martin
  - *Linguistic Structure Prediction*. Noah Smith (no relation, but sometime coauthor)
Thanks