

Natural Language Processing

CSE 40657/60657

Fall 2023

Three Goals



machine
translation



natural
language
interfaces



information
extraction

Three Goals



machine
translation



natural
language
interfaces



information
extraction

Three Goals



machine
translation



natural
language
interfaces



information
extraction

Three Goals



machine
translation

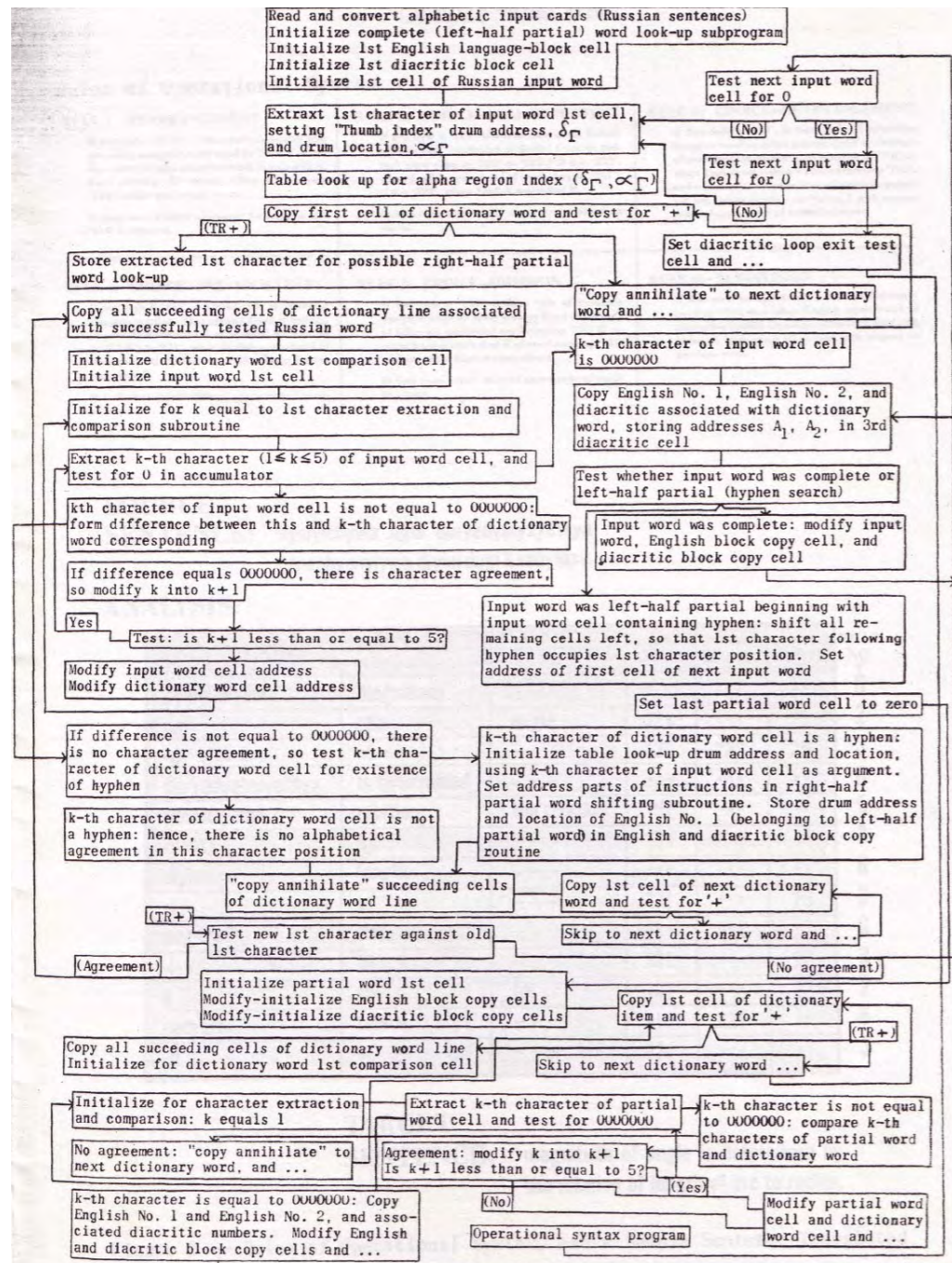


natural
language
interfaces



information
extraction

Machine Translation



DA Please translate from Russian to English: Mi pyeryedayem mislyi posryedstvom ryechyi.

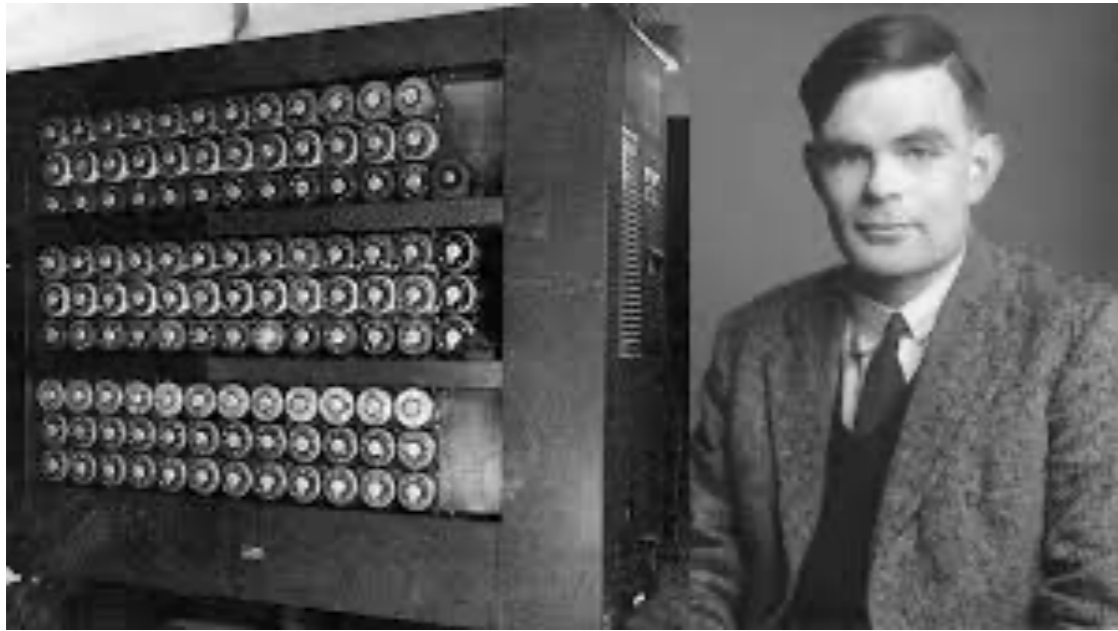
The translation of "Мы передаём мысли посредством речи" from Russian to English is: "We convey thoughts through speech."

DA Please translate from Russian to English: Vyelyichyina ugla opryedyelyayetsya otnoshenyiyem dlyini dugi k radiusu.

The translation of "Величина угла определяется отношением длины дуги к радиусу" from Russian to English is: "The magnitude of the angle is determined by the ratio of the arc length to the radius."

The Georgetown-IBM machine translation experiment, 1954

Natural Language Interfaces



Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764

A: (Pause about 30 seconds and then give as answer) 105621.

Q: Do you play chess ?

A: Yes. .

Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play ?

A: (After a pause of 15 seconds) R-R8 mate.

A. M. Turing, "Computing machinery and intelligence," 1950.



Please write me a sonnet on the subject of the Forth Bridge.



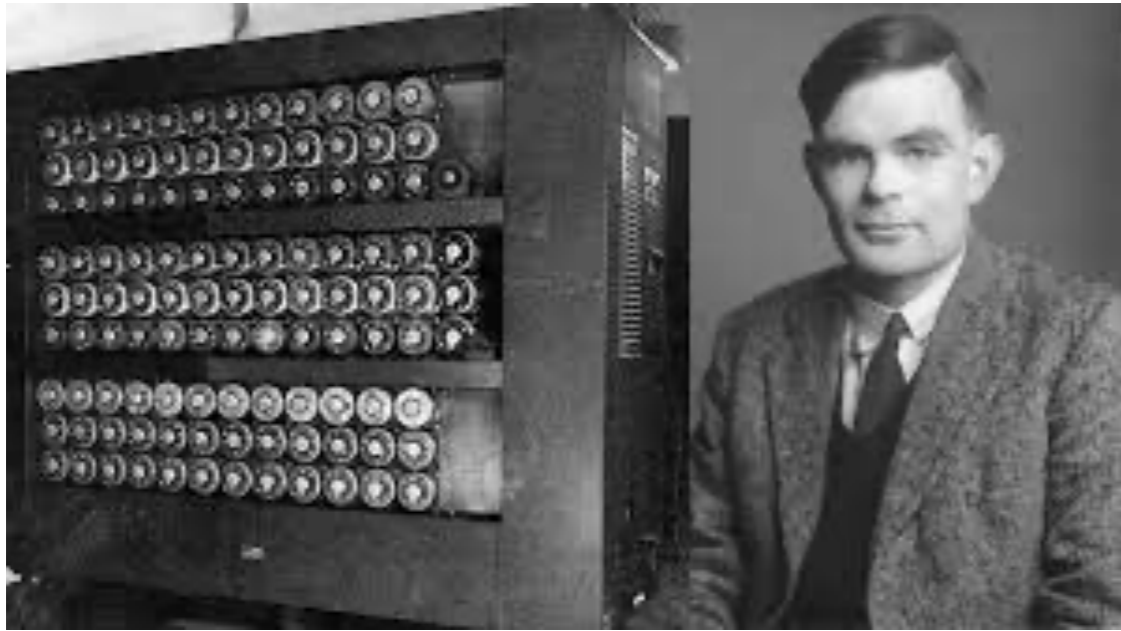
Amidst the rolling waves and salty air,
A marvel stands with strength that ne'er abates,
The Forth Bridge, a symbol grand and rare,
Connecting lands through iron-laden gates.

Its mighty spans in red embrace the sky,
A testament to human skill and might,
A feat of engineering that draws the eye,
A bridge that spans both day and starry night.

For decades now, it's stood with grace untold,
Its towers rise, unyielding 'gainst the tide,
A link 'twixt shores where tales of old are told,
A legacy of progress, far and wide.

So let us marvel at this bridge so grand,
A union of the land and sea and man.

Natural Language Interfaces



Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764

A: (Pause about 30 seconds and then give as answer) 105621.

Q: Do you play chess?

A: Yes.

Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A: (After a pause of 15 seconds) R-R8 mate.

A. M. Turing, "Computing machinery and intelligence," 1950.

DA Do you play chess?



I don't play chess myself, but I'm familiar with the rules and strategies of the game. If you have any questions about chess or need assistance with understanding certain aspects of the game, feel free to ask!



DA

I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?



In the situation you've described, I would play Rb1#. This move would place my rook at R1 and deliver checkmate to your king at K1, ending the game.

Information Extraction



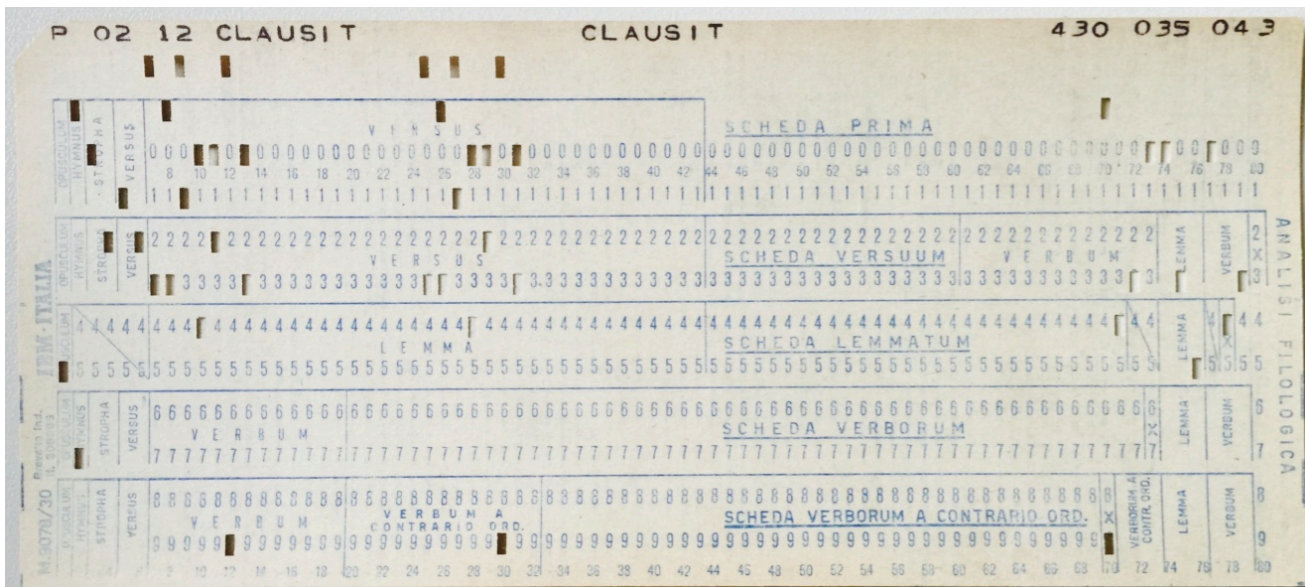
DA

Please explain the concept of presence according to St. Thomas Aquinas.



St. Thomas Aquinas' concept of presence extends beyond the doctrine of the Eucharist and also encompasses the broader philosophical understanding of presence and existence. In his philosophy, presence is closely linked to the metaphysical concepts of substance and accidents.

1. **Substance and Accidents:** Aquinas distinguishes between the substance and accidents of a thing. The substance is the underlying, unchanging reality or essence of an object, while the accidents are the mutable, external characteristics that we perceive through our senses, such as color, taste, size, and shape.
2. **Real and Objective Presence:** Aquinas' philosophy emphasizes the objective reality of things. He believed that things exist in themselves, independently of our perceptions or thoughts about them. In this context,



Roberto Busa, SJ, and the *Index Thomisticus*

However...



Please write a sentence that starts and ends with the same word.



"Time" is both a constant companion and an enigmatic force, shaping the very existence of our universe.



Does "0110101" have an even or odd number of "1"s?



The binary number "0110101" has an odd number of "1"s. There are a total of 4 "1"s in the sequence.



Please insert "freakin" into the word "fantastic."



Sure, the modified word would be "fanfreakintastic."



Why can't it be "fantasfreakintic"?



You're absolutely right, "fantasfreakintic" would also be a valid way to insert "freakin" into the word "fantastic." My previous response was just one possible option, but your suggestion works perfectly fine as well!



Please insert "freakin" into the word "computational."



Certainly! The modified word would be "comfreakintational."

However...

DA

What is wrong with the sentence "Who was John wondering whether could eat peanuts?"



The sentence "Who was John wondering whether could eat peanuts?" is a bit awkward and unclear. The main issue is with the placement of "whether" and the lack of a clear subject for the action. To make the sentence clearer, you could rephrase it as:

"John was wondering whether someone could eat peanuts."

DA

In response to the question, "Who would like ice cream," would you say, "I would!" or "I'd!"?



Both "I would!" and "I'd!" are correct responses to the question "Who would like ice cream?" They both convey the idea that you would like ice cream. "I'd" is a contraction of "I would," so either response is appropriate and grammatically accurate. The choice between the two depends on your personal preference and the level of informality you want to convey.

Two Big Ideas

Linguistics

Machine
Learning

1950

1970

1990

2010

Two Big Ideas

Linguistics

Machine
Learning

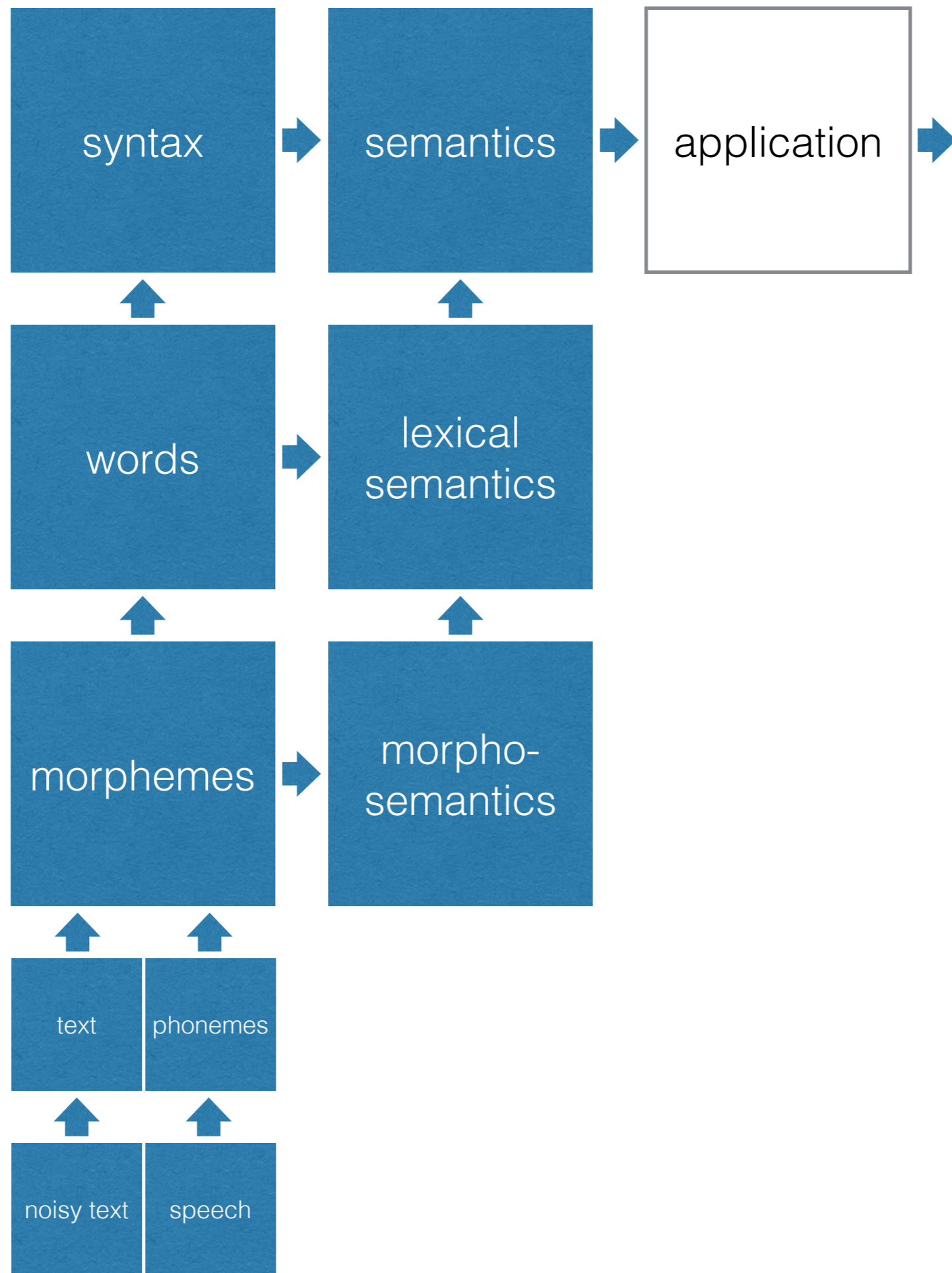
1950

1970

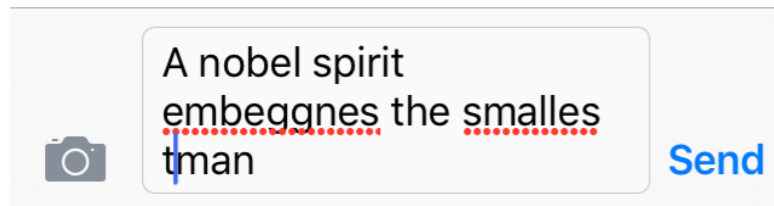
1990

2010

ALPAC report, 1966

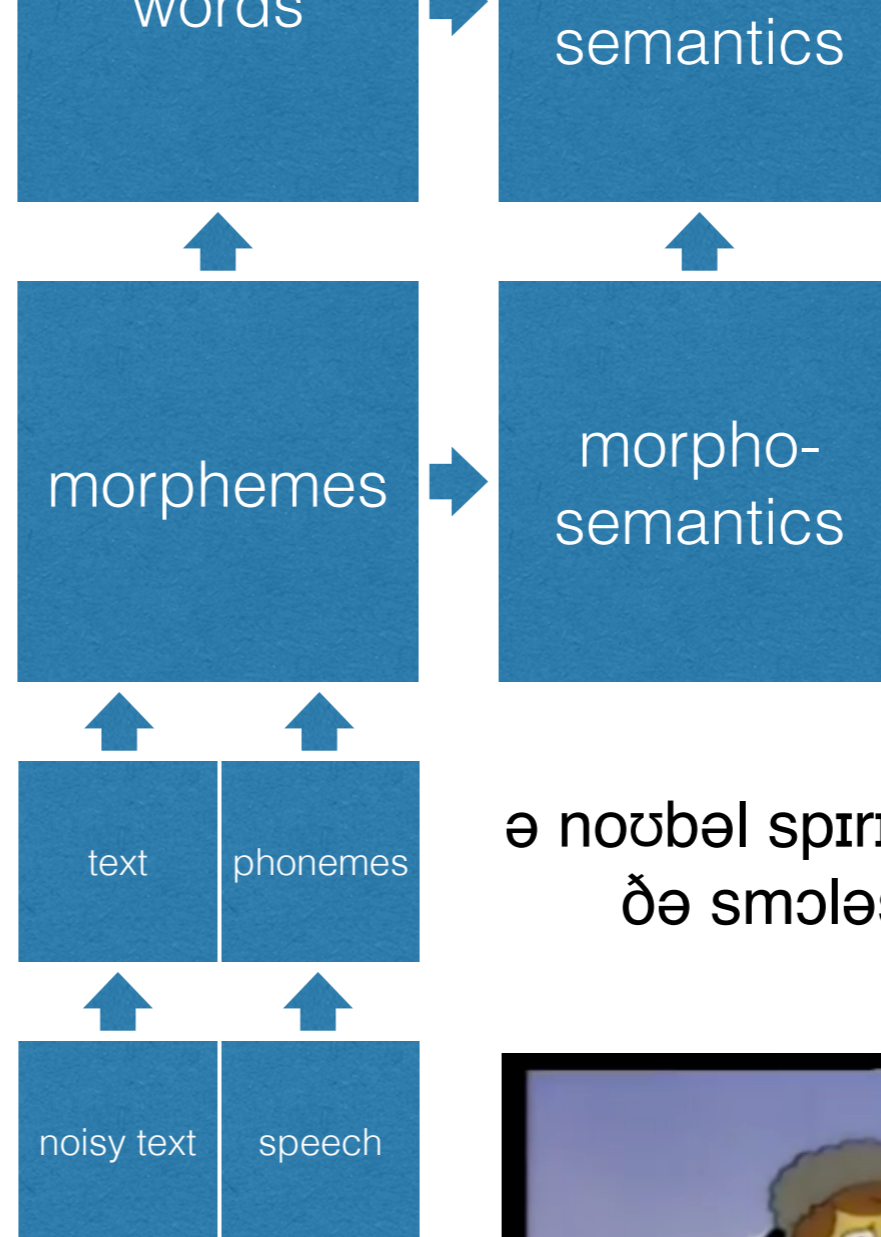


a noble spirit embiggens
the smallest man



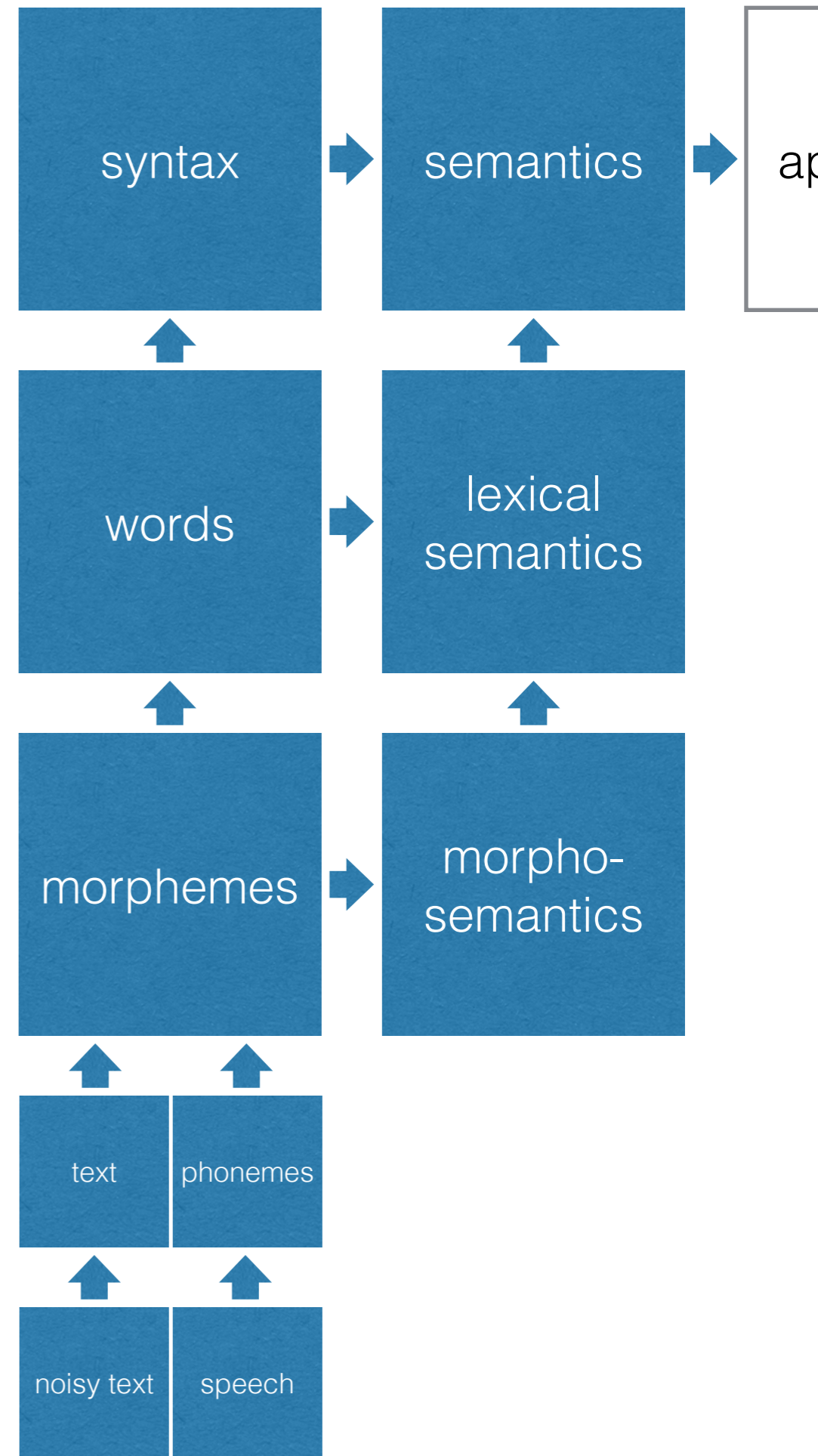
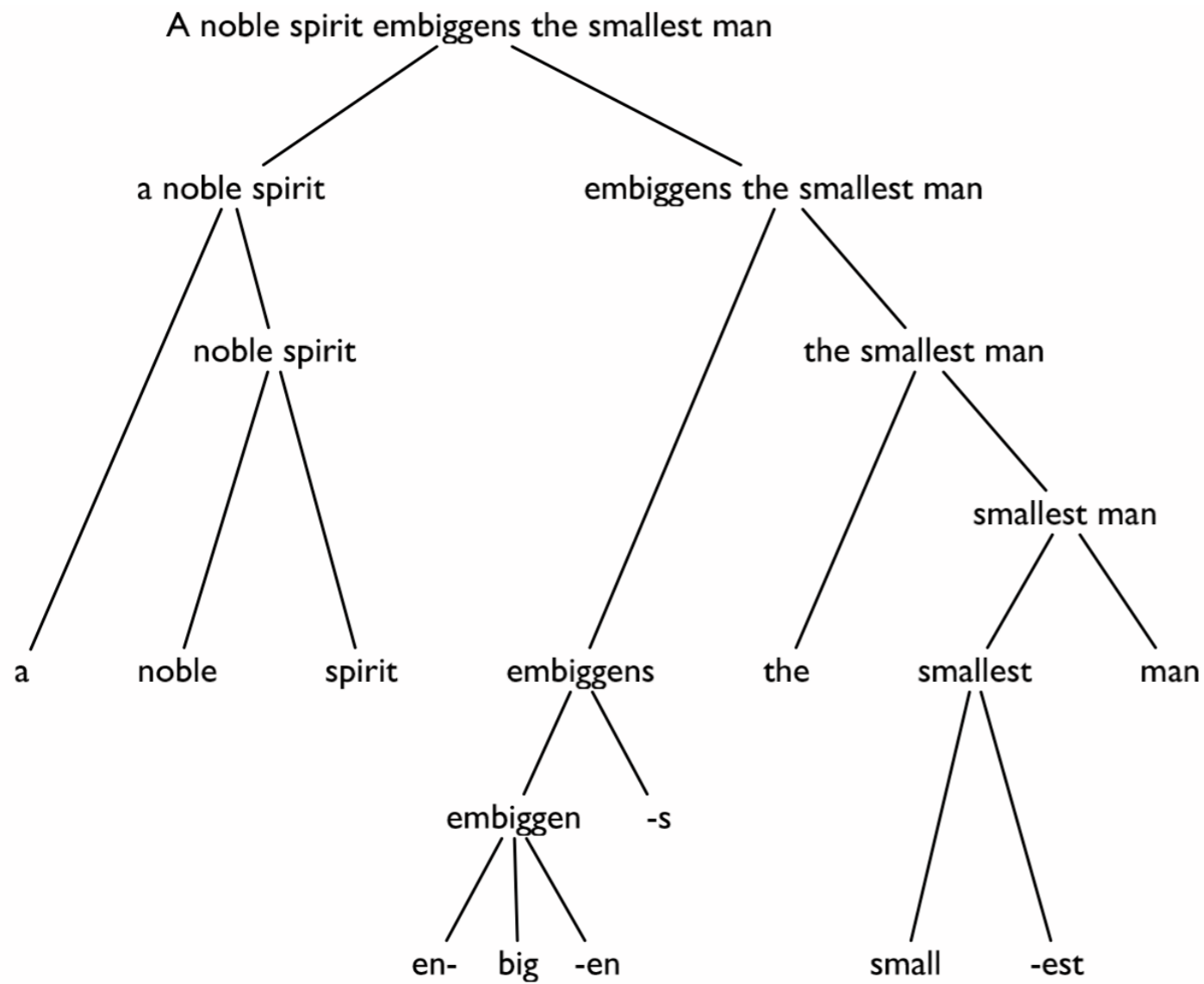
A noble spirit embiggens the smallest man.

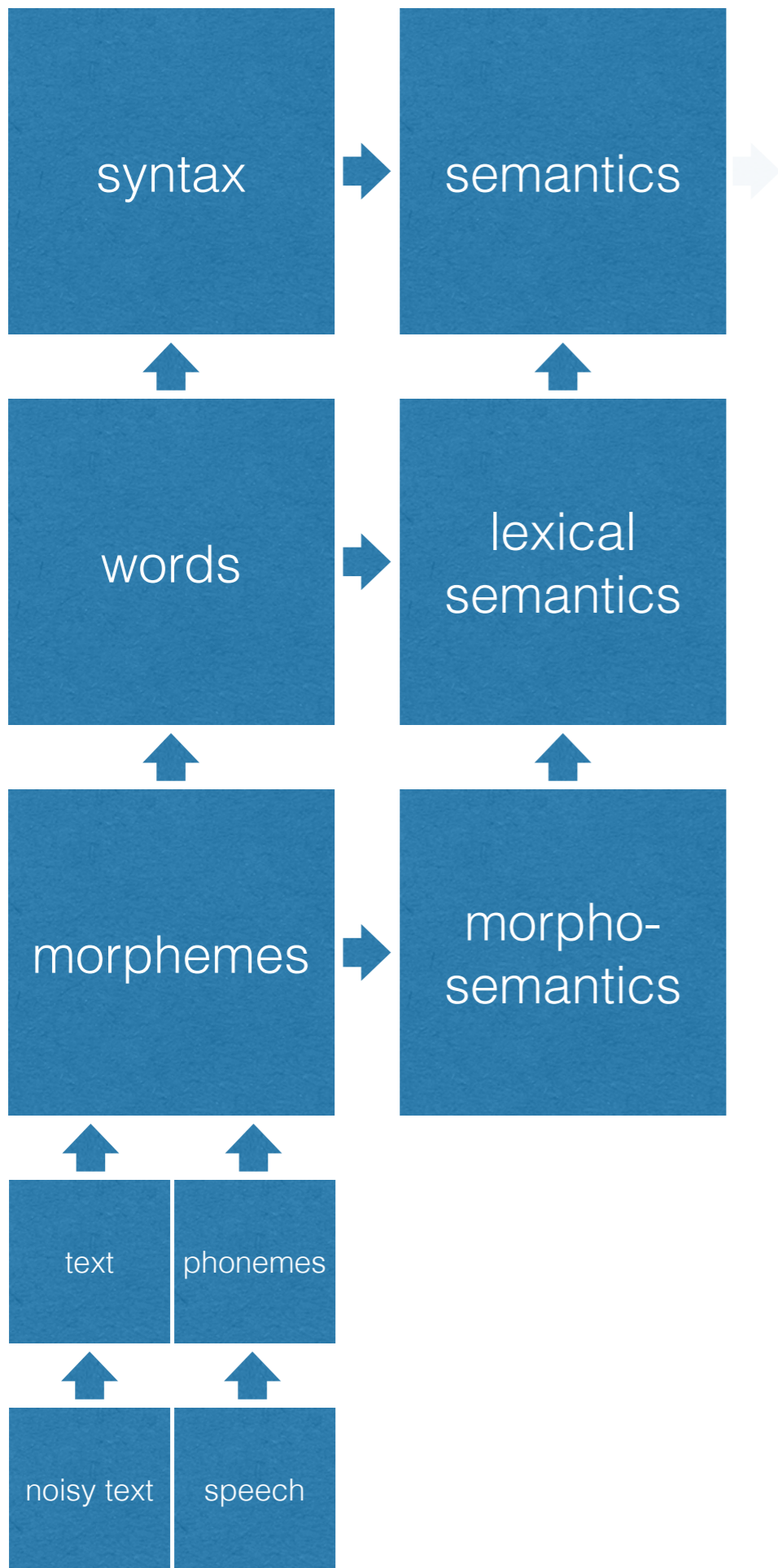
A noble spirit embiggens the smallest man.



ə noubəl spɪrɪt ɛmbɪgənz
ðə smɔləst mæn







application

$\exists s.\exists m.\text{spirit}(s)\wedge\text{noble}(s)\wedge\text{man}(m)\wedge$
 $\forall m'.\text{smaller}(m, m')\wedge\text{enlarge}(s,m)$

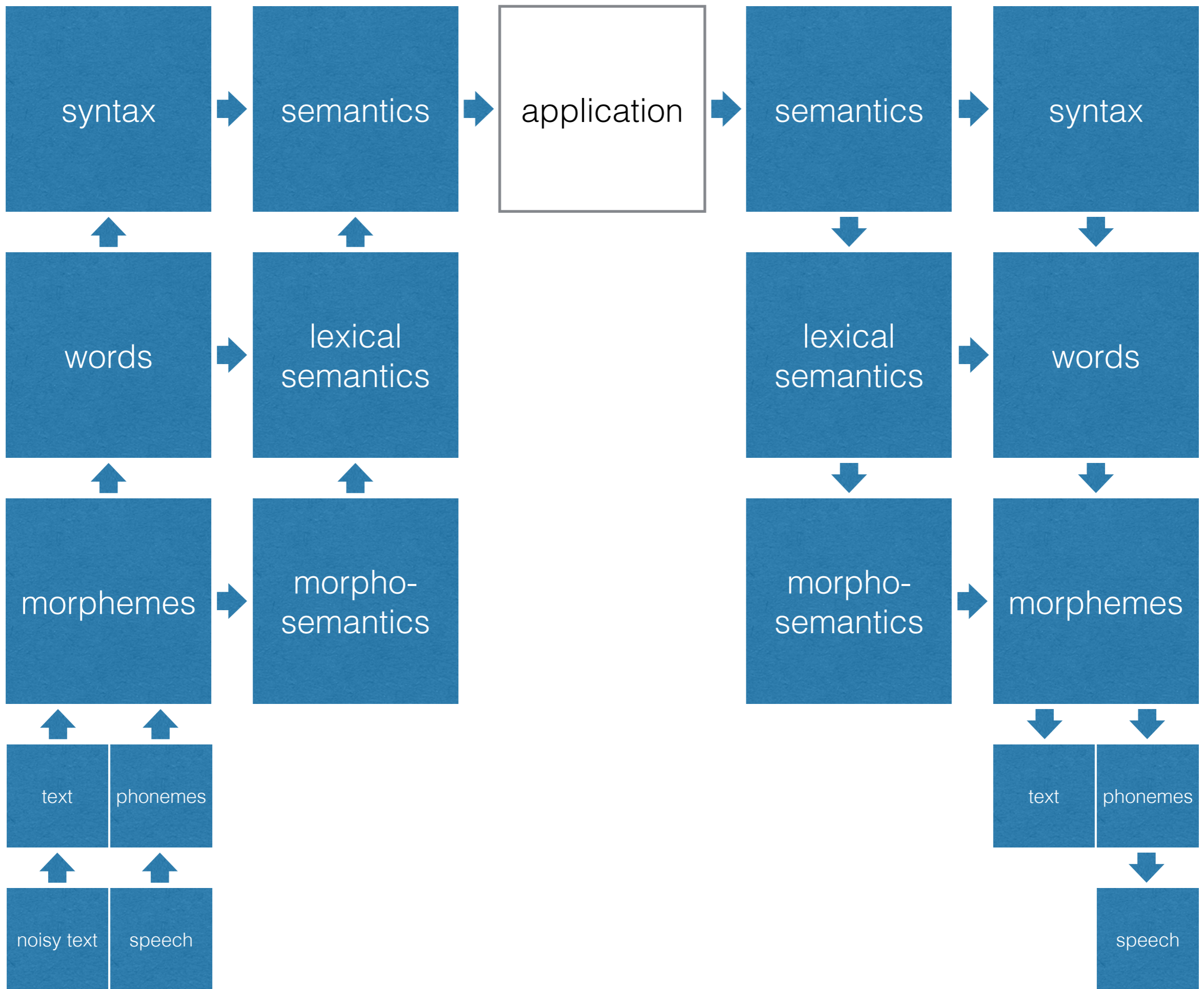
Predicate: *enlarge*

Roleset id: *enlarge.01* , *make bigger*, Source: , vncls: , framnet:

enlarge.01: ENLARGE-V NOTES: Preposition 'by' takes precedence over EXT tag. (from enlarge.01-v) ENLARGED-J NOTES: Based on Thyme and Sharp data. Comparison to enlarge.01. Enlarge.01 is a member of VN class other_cos-45.4. Framed by Katie. (from enlarged.01-j)

Roles:

- Arg0-PAG:** *causer of biggenment* (vnrole: 45.4-agent)
- Arg1-PPT:** *thing getting bigger* (vnrole: 45.4-patient)
- Arg2-EXT:** *EXT*
- Arg3-DIR:** *start point*
- Arg4-GOL:** *end point*



Two Big Ideas

Linguistics

Machine
Learning

1950

1970

1990

2010

Two Big Ideas

Machine
Learning

Linguistics

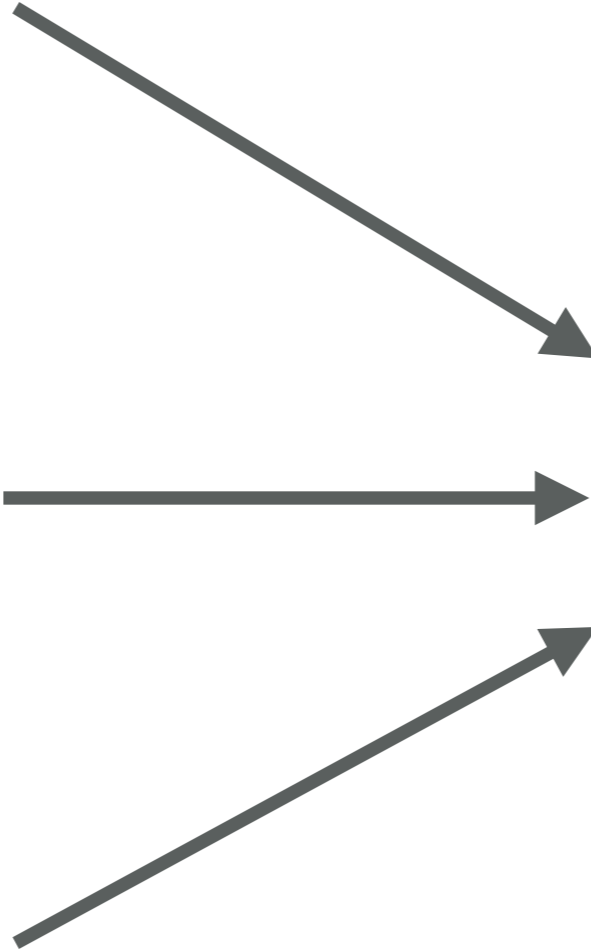
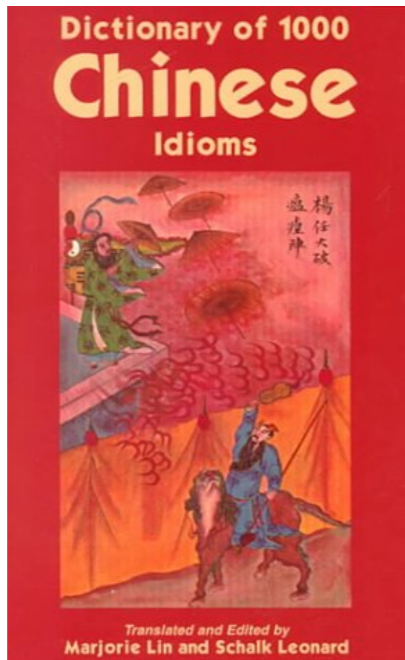
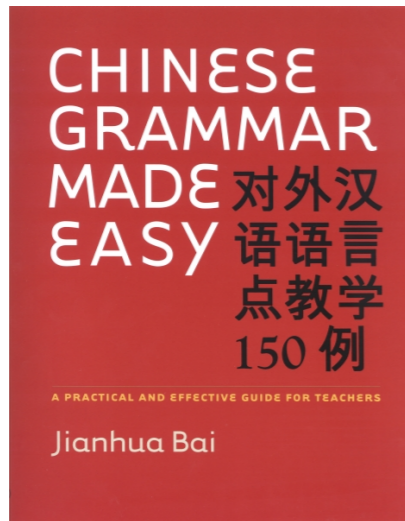
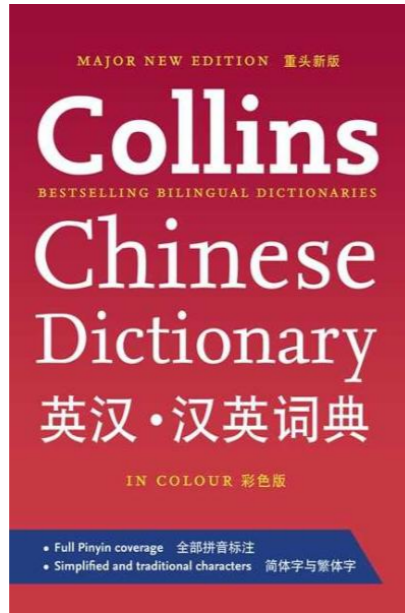
1950

1970

1990

2010

DARPA Speech and Natural
Language Workshop, 1989



口在牙头加以南达到飓风强度标准，并在接下来数天。11月17日，莱尼在加勒比海东北部快速强化，在美海域达到每小时250公里的最大风速。飓风在途经背风洋开放海域上空消散。行经小安的列斯群岛前，莱尼。狂风暴雨导致波多黎各东南部的农作物严重受损。出现大范围洪灾和侵蚀，但破坏只达到“中等程度”。美美元的损失。降雨量最高的是圣马丁岛法国领地的警0套物业被毁。附近的安提瓜和巴布达也有1人丧生，也受到这场风暴的严重破坏，大浪导致多个城镇和首损失，其名称“莱尼”因此从大西洋风暴命名名单上退命名时使用。第十世班禅额尔德尼是藏传佛教格鲁派仅次于达赖喇嘛。1949年6月3日在时任国民政府代总武继任。同年9月，解放军占领西宁后，十世班禅主动十世班禅、十四世达赖喇嘛与中央政府签订《关于和区骚乱后，十世班禅选择留在西藏，接受了中共任命对中共在藏区实施民主改革中产生问题的不满，十世的西藏政策进行了严厉的挞伐。1964年，十世班禅被“最大的反动农奴主之一”并被剥夺了一切职务。1966班禅集团”受到迫害入狱。直到1978年，在得到时任中十世班禅得以出狱。出狱后，十世班禅恢复职务，并、语言、习俗的重建与改革，得到了许多藏族人的爱部的城市，是威斯特法伦地区的文化中心。亦是明斯是明斯特反叛的发生地，此地也是1648年标志着三十。今天，该城以德国的自行车之都闻名。1915年明斯



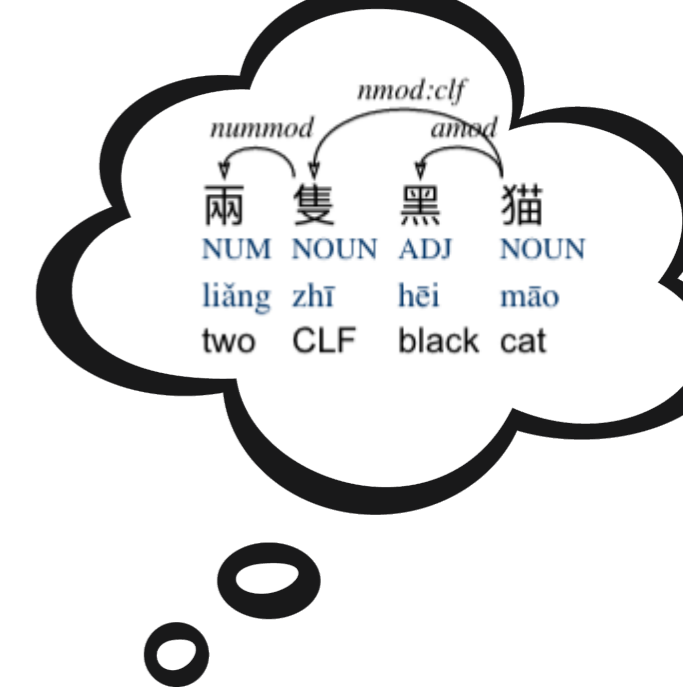
(IP (NP-SBJ (NN 中资))
 (VP (ADVP (AD 已))
 (VP (VV 成为)
 (NP-OBJ (NP-PN (NR 澳门))
 (CP (WHNP-1 (-NONE- *OP*))
 (CP (IP (NP-SBJ (-NONE- *T*-1))
 (VP (ADVP (AD 最))
 (VP (VA 大))))))
 (DEC 的)))
 (ADJP (JJ 外来))
 (NP (NN 投资者))))))
 (PU 。))

(IP (NP-SBJ (DP (DT 这)
 (CLP (M 座)))
 (NP (NN 楼)))
 (VP (QP-PRD (CD 十八)
 (CLP (M 英尺))))))

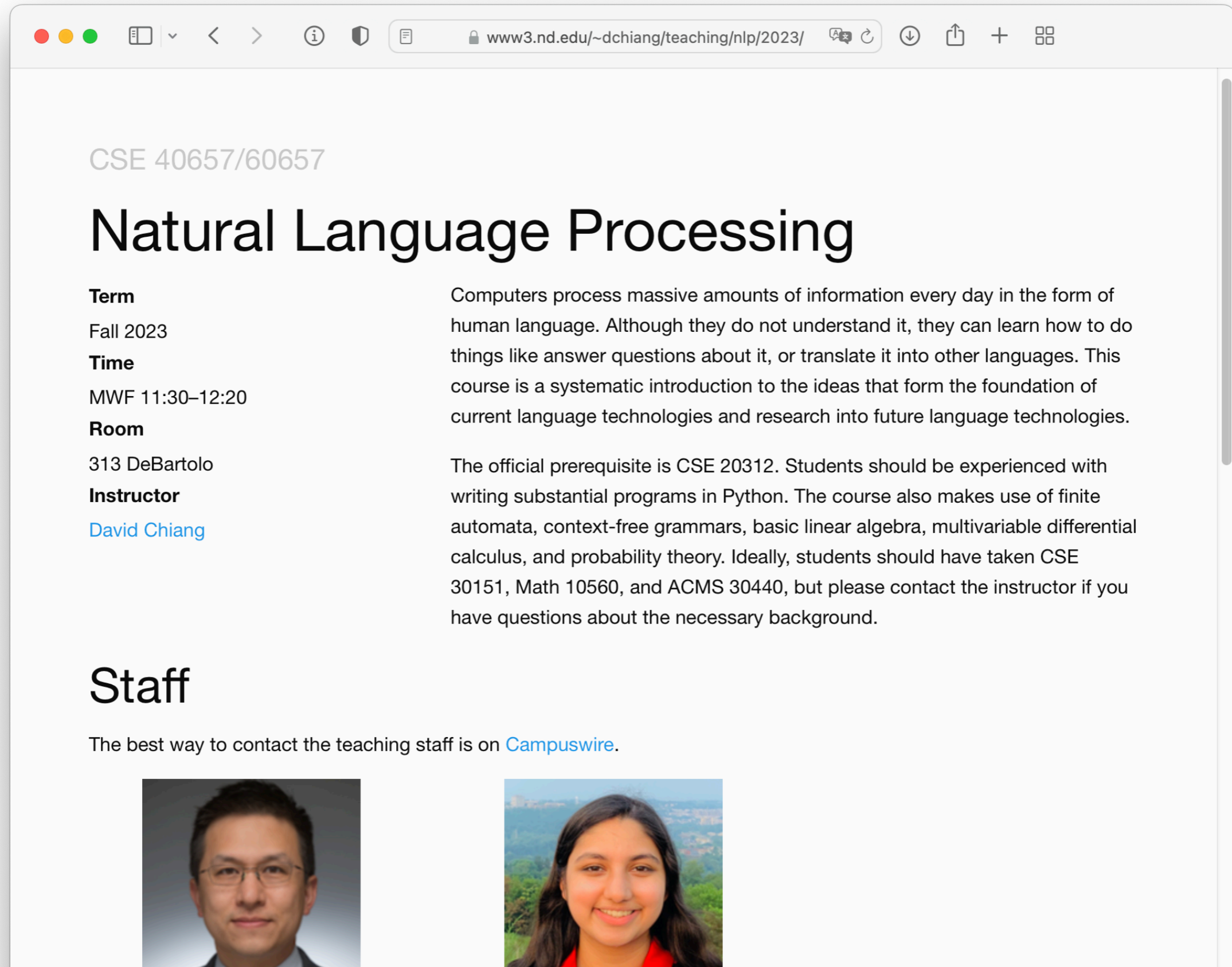
(VP (VV 制定)
 (AS 了)
 (NP-OBJ (IP-APP (NP-SBJ (-NONE- *pro*))
 (VP (VP (VV 引进)
 (NP-OBJ (NN 外资)))
 (PU 、)
 (VP (VV 加强)
 (NP-OBJ (ADJP (JJ 横向))
 (NP (NN 经济)
 (NN 联合))))))
 (CC 和)
 (VP (PP (P 对)
 (NP (NN 外)))
 (VP (VV 下放)
 (NP-OBJ (NN 权))))))
 (QP (CD 三)
 (CLP (M 个)))
 (NP (NN 文件))))



口在牙头加以南达到飓风强度标准，并在接下来数天。11月17日，莱尼在加勒比海东北部快速强化，在美海域达到每小时250公里的最大风速。飓风在途经背风洋开放海域上空消散。行经小安的列斯群岛前，莱尼。狂风暴雨导致波多黎各东南部的农作物严重受损。出现大范围洪灾和侵蚀，但破坏只达到“中等程度”。美美元的损失。降雨量最高的是圣马丁岛法国领地的警0套物业被毁。附近的安提瓜和巴布达也有1人丧生，也受到这场风暴的严重破坏，大浪导致多个城镇和首损失，其名称“莱尼”因此从大西洋风暴命名名单上退命名时使用。第十世班禅额尔德尼是藏传佛教格鲁派仅次于达赖喇嘛。1949年6月3日在时任国民政府代总式继任。同年9月，解放军占领西宁后，十世班禅主动十世班禅、十四世达赖喇嘛与中央政府签订《关于和区骚乱后，十世班禅选择留在西藏，接受了中共任命对中共在藏区实施民主改革中产生问题的不满，十世的西藏政策进行了严厉的挞伐。1964年，十世班禅被“最大的反动农奴主之一”并被剥夺了一切职务。1966班禅集团”受到迫害入狱。直到1978年，在得到时任中十世班禅得以出狱。出狱后，十世班禅恢复职务，并、语言、习俗的重建与改革，得到了许多藏族人的爱部的城市，是威斯特法伦地区的文化中心。亦是明斯是明斯特反叛的发生地，此地也是1648年标志着三十。今天，该城以德国的自行车之都闻名。1915年明斯



nd.edu/~dchiang/teaching/nlp



The screenshot shows a web browser window with the address bar containing the URL `www3.nd.edu/~dchiang/teaching/nlp/2023/`. The page content includes the course ID `CSE 40657/60657`, the title `Natural Language Processing`, and a list of course details: `Term: Fall 2023`, `Time: MWF 11:30–12:20`, `Room: 313 DeBartolo`, and `Instructor: David Chiang`. A paragraph describes the course as a systematic introduction to NLP, and another paragraph lists prerequisites: `CSE 20312`, `CSE 30151`, `Math 10560`, and `ACMS 30440`. A `Staff` section follows, with a note that the best way to contact the teaching staff is on `Campuswire`. At the bottom, there are two portrait photos: one of a man with glasses and one of a woman with long dark hair.

CSE 40657/60657

Natural Language Processing

Term
Fall 2023

Time
MWF 11:30–12:20

Room
313 DeBartolo

Instructor
[David Chiang](#)

Computers process massive amounts of information every day in the form of human language. Although they do not understand it, they can learn how to do things like answer questions about it, or translate it into other languages. This course is a systematic introduction to the ideas that form the foundation of current language technologies and research into future language technologies.

The official prerequisite is CSE 20312. Students should be experienced with writing substantial programs in Python. The course also makes use of finite automata, context-free grammars, basic linear algebra, multivariable differential calculus, and probability theory. Ideally, students should have taken CSE 30151, Math 10560, and ACMS 30440, but please contact the instructor if you have questions about the necessary background.

Staff

The best way to contact the teaching staff is on [Campuswire](#).

