Introduction
CSCI 394: Computational Linguistics

CSCI 394

August 29, 2013

CL in a Google search

CL in a Google search

CSCI 394   Introduction

Grammar checking in Word

There is many examples of computational linguistics around us.
Neither of these is right.
Neither of these are right.
Many people use incorrect grammar.
A people stands or falls on its use of right grammar.

Help

A people stand
Person stands

Ignore
About this Sentence

Grammar...
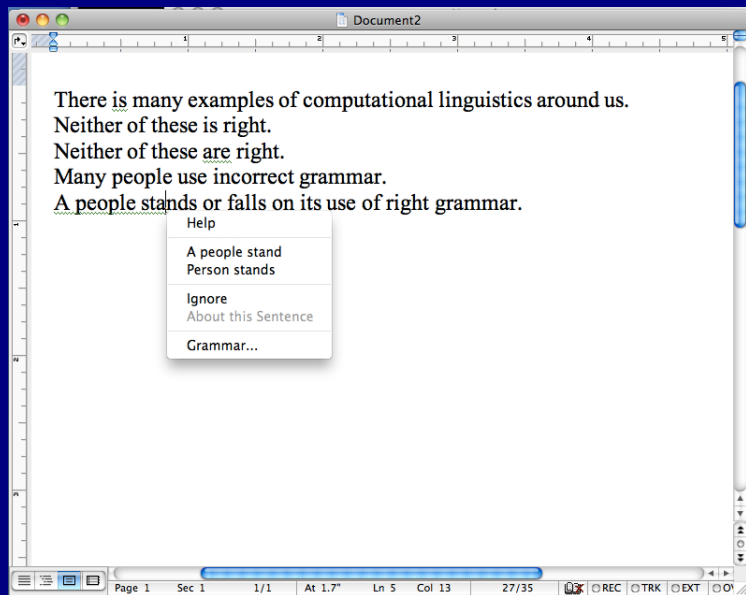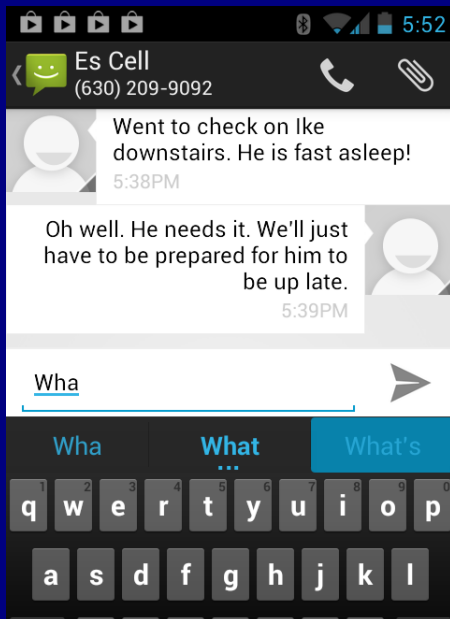
Text completion
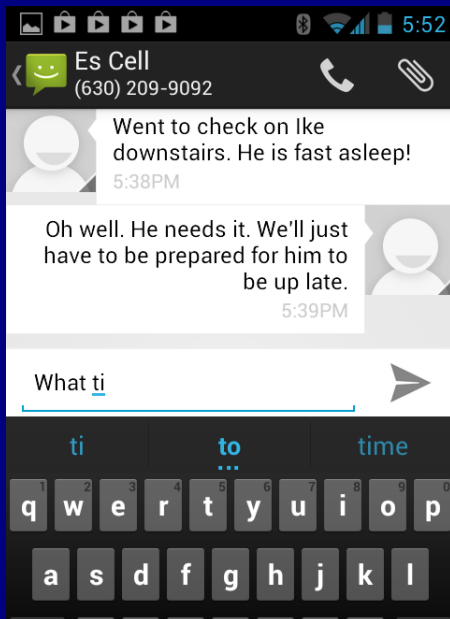
Text completion

Text completion

Text completion

Text completion
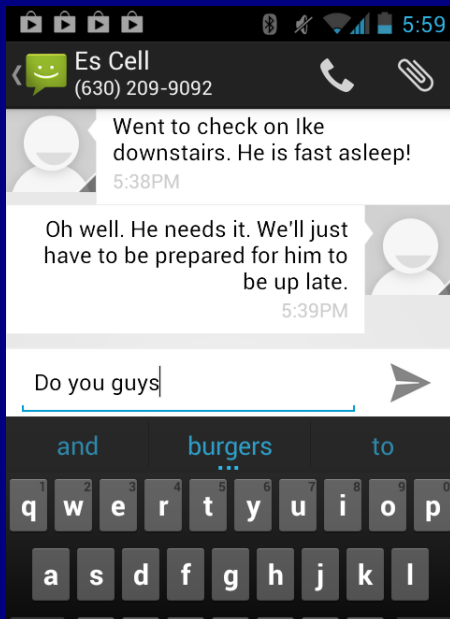
IBM Watson

Source: CNN.com

Apple Siri

Source: apple.com

C4 | Saturday/Sunday, May 4 - 5, 2013   * * * *   THE WALL STREET JOURNAL.

IDEAS MARKET

WEEK IN WORDS: ERIN McKEAN

A field guide to unusual words in this week's Wall Street Journal.

Tasty Butterfly
papalo

In the summer, she chops papalo up for homemade salsas and says it tastes like "cilantro on steroids."

—"Beyond Basic Herbs: Lesser-Known Greens Get Trendy," Personal Journal, May 1

Papalo comes from the word for "butterfly" in Nahuatl, the language of the Nahua people, many of whom live in central Mexico. It's also known as summer cilantro and Bolivian coriander (although botanically, it's unrelated to coriander).

A Cousin of Dozen
duodecad

The Hill reports that the duodecad... last week blamed global warming for "drought and reduced agricultural output," which they claim "can be particularly harmful for women."
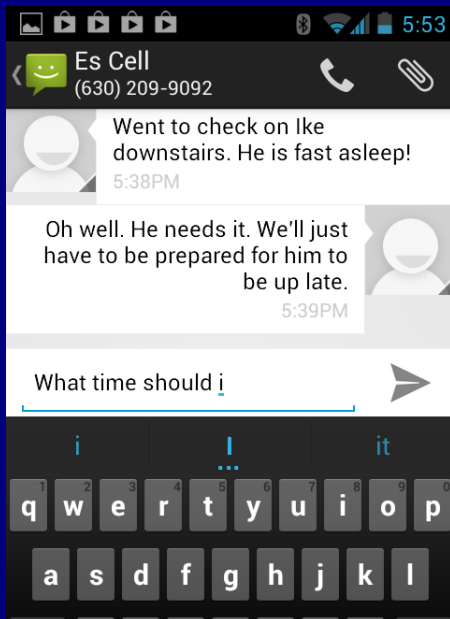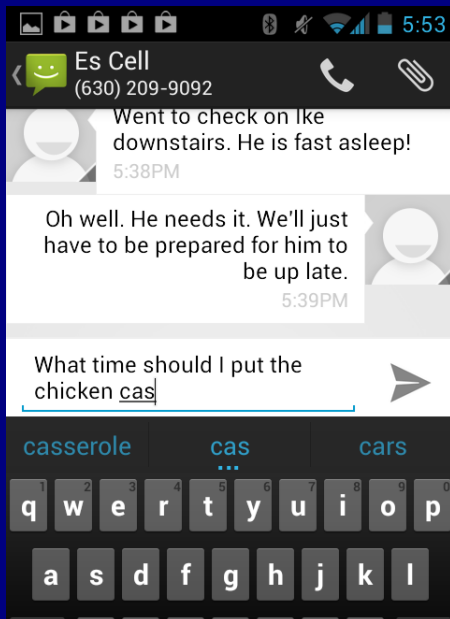
SENTIMENT TRACKER: A COMPUTATIONAL ANALYSIS OF THE CONVERSATION ON SOCIAL NETWORKS

Out of 3,400 posts

THE ORACLE OF CYBERSPACE?
The online buzz about Warren Buffett joining Twitter.

39% EXCITED

'You are a wonderful human being, Warren. Thank you for all you have done for the world. And welcome to Twitter!'

'Nice to have the "Oracle of Omaha" on Twitter. Looking forward to his insight.'

'You are a hero to me! So glad you set up an account.'

FIRST TWEET: '@Warren Buffett: Warren is in the house.'

Associated Press

1% DON'T CARE

'Following the richest man on Twitter won't help you get rich, people.'

'I guess people are starved for insight.'

57% AMAZED AT NUMBER OF FOLLOWERS

'Warren Buffett is racking up followers faster than you can shake a fist full of money.'

'Every time I refresh Buffett's profile, he's gained 200 followers. Wow.'

'His Twitter following is growing as fast as his bank account.'

3% JOKES

'What's next? The Pope joins Pinterest?'

'Justin Bieber still has more followers.'

'I already feel wealthier and more intelligent.'

'Now he can mess with the stock market in real time.'

'The average net worth of Twitter users just jumped.'

Machine translation

NIV revision
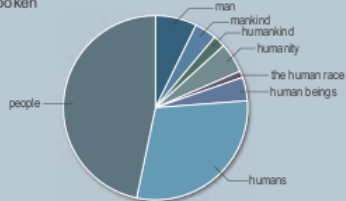

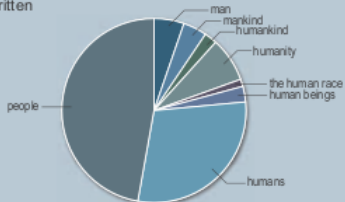Mankind, man and synonyms in (2005-2009)
General written
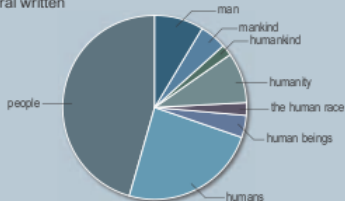
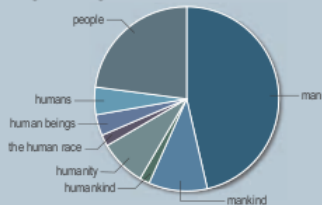Mankind, man and synonyms in (2005-2009)
US spoken

Mankind, man and synonyms in (2005-2009)
US written

Mankind, man and synonyms in (2005-2009)
Evangelical English

Source: Collins Dictionaries, "The Development and Use of Gender Language in Contemporary English"

Tokens, types, and lemmas

I rose to saw off the still rose that I saw still grew near the still.
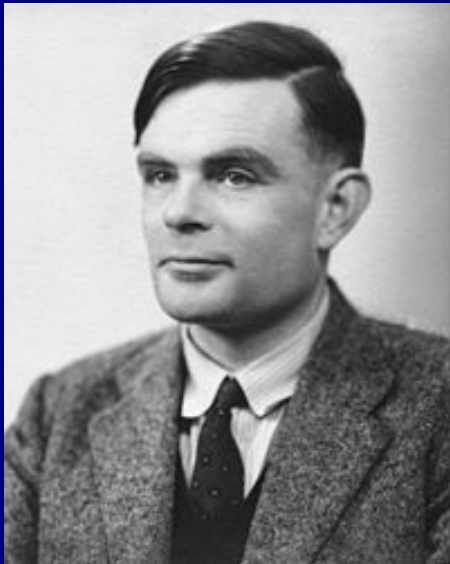
16 tokens.

Types occurring once: to, off, that, grew, near.

Types occurring twice: I, rose, saw, the.

Type occurring three times: still

Distinct lemmas: I, rise, to, saw (verb-cut), off, the, still (verb), rose (noun), that, see, still (adverb), grow, near, still (noun).
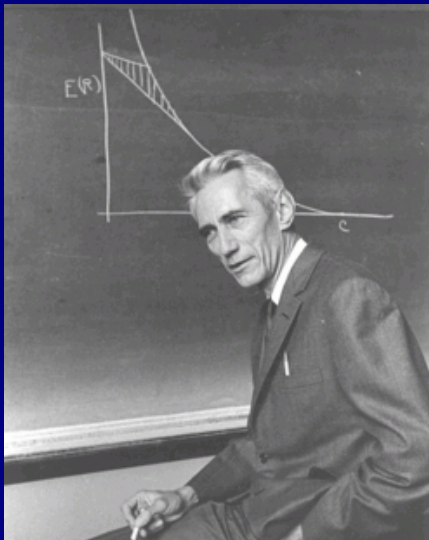
Alan Turing, 1912–1954

Source: Wikipedia

- World War II cryptanalysis
- Theory of computation (abstract machines deciding languages)
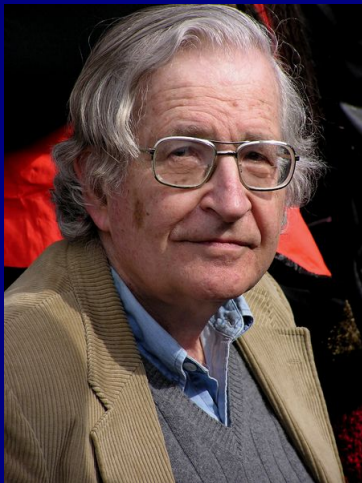- The Turing test of machine intelligence

Claude Shannon, 1916–2001

Source: nyu.edu

- World War II cryptanalysis
- Information theory
- Noisy channel mode
- Entropy

Noam Chomsky, 1928–

Source: Wikipedia (Duncan Rawlinson)

- Formal linguistics
- Phrase-structural, generative grammars
- Context-free grammars
- Language heirarchy

Source: mit.edu

- ELIZA conversational agent
- Social and ethical consequences of technology