

AIRTA: An Automatic Interdisciplinary Research Topic Advisor

- Where are We and Where do We Go -

Michael Paul† and Roxana Girju‡

Departments of Computer Science(† ‡) and Linguistics (‡), Beckman Institute († ‡)

University of Illinois at Urbana-Champaign

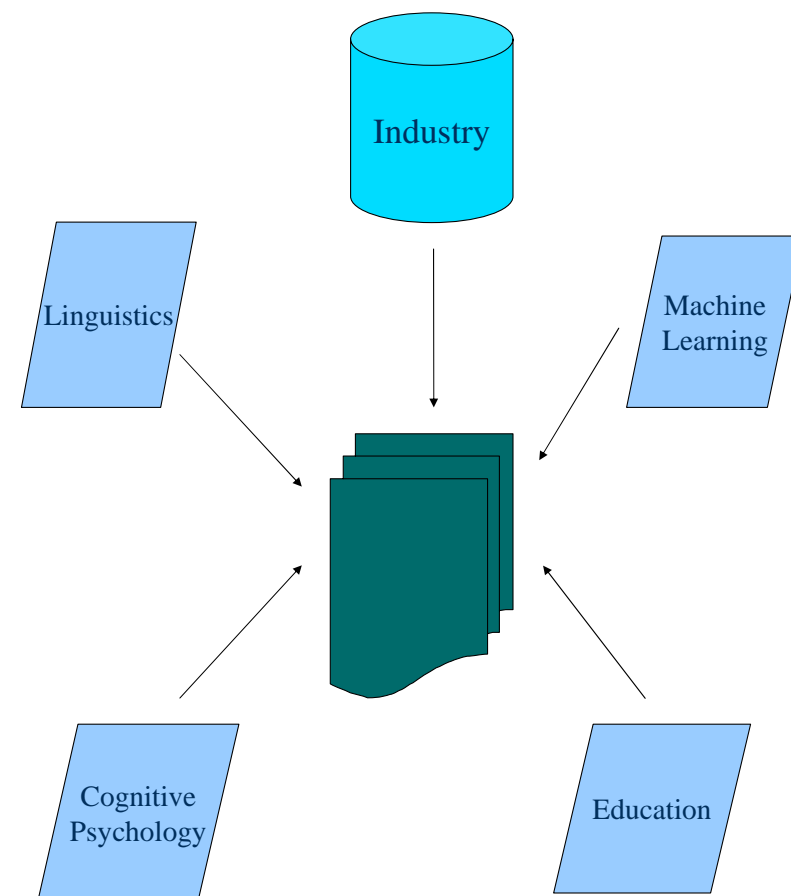
{mjpaul2, girju}@illinois.edu

Introduction

We believe that like other disciplines, computational linguistics will drastically benefit from an inter-disciplinary perspective.

Our tool is designed to foster interdisciplinary research in order to make breakthrough predictions for future directions.

This is accomplished by analysing trends within and across relevant fields and then automatically suggesting new research directions and topics.



Some fields motivating research in computational linguistics

Back End

We currently have a database with:

- 4,700 papers from **computational linguistics** conferences
- 2,300 papers from **linguistics** journals
- 1,700 papers from **education/educational psychology** journals

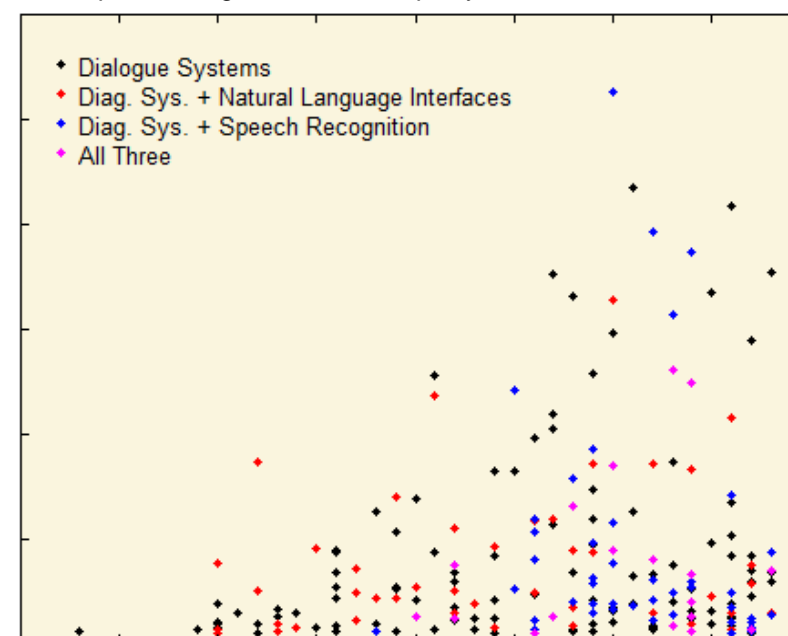
We will enlarge our corpus as we continue to work on this project.

Classification

We categorized these papers mostly using Latent Dirichlet Allocation (LDA) with words from titles, abstracts, and full text when available.

Lexical Semantics	lexical entries semantic word idioms words lexicon
Morphology	morphological word morphology lexical level forms
MT Evaluation	evaluation score human scores sentence automatic
Multimodal NLP	multimodal speech gesture user language input
Named Entities	entity names named entities ne information person

A sample of categories and the top keywords associated with them



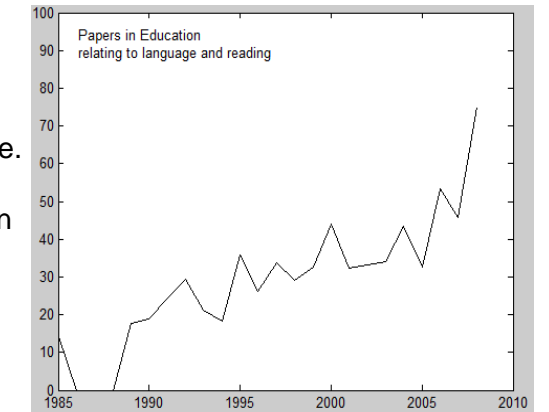
Each dot represents a paper in the "Dialogue Systems" category. The coloring shows how papers can span multiple categories.

Trends/Analysis

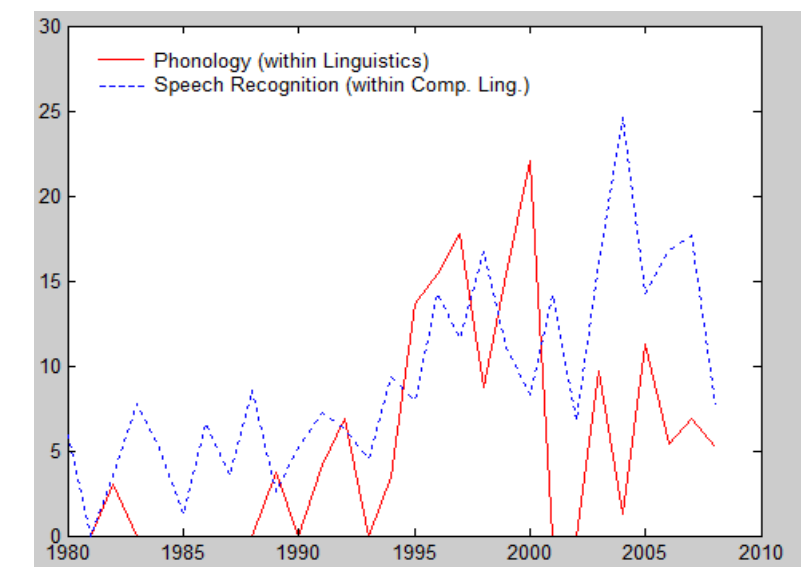
Because our data is categorized and labelled by year, we can see how research in certain fields rises and declines over time.

We can use this information to gauge which topics are important and which areas are saturated.

We also look for correlations in trends in similar fields across different disciplines.



Language-related topics comprise the bulk of research in education and are steadily increasing in prominence.



Next Step

The next phase of this project (the final goal) will be to generate new topics. The key is to discover topics that are important in one discipline but have been studied little in another. These suggestions will be useful to professionals who would like to engage in research discussions with other parties, but who are not familiar with those areas. It will be beneficial to students looking for novel research topics.