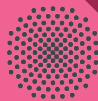# Distributional Information: A Powerful Cue for Acquiring Syntactic Categories

Levindo Gabriel Taschetto Neto
22.11.2017

Probabilistic models of language and cognition

Universität Stuttgart

Agenda

1. Introduction

- Distributional information is a potentially important source of data for identifying the syntactic categories of words.
- Distributional information provides a powerful cue for acquiring syntactic categories.

2. Learning Syntactic Categories' Problem

- A completely unconstrained search with <u>n items</u> and <u>m syntactic categories</u> (assuming, for simplicity, that each item has a single syntactic category), would involve considering m in the power of n possible mappings.

2. Learning Syntactic Categories' Problem

- A completely unconstrained search with <u>n items</u> and <u>m syntactic categories</u> (assuming, for simplicity, that each item has a single syntactic category), would involve considering m in the power of n possible mappings.

Example:

- Given 3 syntactic categories and 15 different items.

2. Learning Syntactic Categories' Problem

- A completely unconstrained search with <u>n items</u> and <u>m syntactic categories</u> (assuming, for simplicity, that each item has a single syntactic category), would involve considering m in the power of n possible mappings.

Example:

- Given 3 syntactic categories and 15 different items.
  n = 15
  m = 3

2. Learning Syntactic Categories' Problem

- A completely unconstrained search with <u>*n* items</u> and <u>m syntactic categories</u> (assuming, for simplicity, that each item has a single syntactic category), would involve considering *m* in the power of *n* possible mappings.
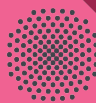
Example:

- Given 3 syntactic categories and 15 different items.
  n = 15
  m = 3
  Possible mappings = 3^15 = 14.348.907

3. Available Information

Based on:

3. Available Information

Based on:

1. Distributional analysis of linguistic input.

3. Available Information

Based on:

1. Distributional analysis of linguistic input.
2. Relation of the linguistic input to the situation.

3. Available Information

Based on:

1. Distributional analysis of linguistic input.
2. Relation of the linguistic input to the situation.
3. Phonological cues to syntactic category.

3. Available Information

Based on:

1. Distributional analysis of linguistic input.
2. Relation of the linguistic input to the situation.
3. Phonological cues to syntactic category.
4. Analysis of prosody.

3. Available Information

Based on:

1. Distributional analysis of linguistic input.
2. Relation of the linguistic input to the situation.
3. Phonological cues to syntactic category.
4. Analysis of prosody.
5. Natural knowledge of syntactic categories.
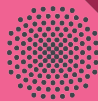
3.1 Distributional analysis of linguistic input

- Information about the linguistic context in which a word occurs.

3.1 Distributional analysis of linguistic input

- Information about the linguistic context in which a word occurs.
- Words of the same category tend to have a large number of distributional regularities in common can be used as a cue to syntactic category.

3.1 Distributional analysis of linguistic input

- Information about the linguistic context in which a word occurs.
- Words of the same category tend to have a large number of distributional regularities in common can be used as a cue to syntactic category.

Example (Maratsos and Chalkley, 1980)

Words which take the suffix *-ed* typically also take the suffix *-s*, and are verbs. Also, words which take the suffix *-s*, but <u>not</u> the suffix *-ed*, are typically count-nouns.
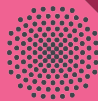
3.2 Relation of the linguistic input to the situation

- A mechanism for the initial classification of words makes use of a correlation between prior semantic categories (such as object and action) in terms of which the child already perceives the world and syntactic categories.

3.3 Phonological cues to syntactic category

- Myriad regularities between the phonology of words and their syntactic categories may be utilized in order to acquire these categories (Kelly, 1992).
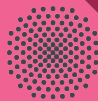
3.3 Phonological cues to syntactic category

- Myriad regularities between the phonology of words and their syntactic categories may be utilized in order to acquire these categories (Kelly, 1992).

  Example:
  English <u>disyllabic nouns</u> tend to have stress on the <u>initial syllable</u>, while <u>verbs</u> have <u>final syllable</u> stress.

3.4 Analysis of prosody

- Learners exploit the mutual predictability between the syntactic phrasing of a sentence, and the way it is said (Morgan and Newport, 1981).

3.5 Natural knowledge of syntactic categories

1. Learning mechanisms that exploit information of any kind in the input may be innately specified.

3.5 Natural knowledge of syntactic categories

1. Learning mechanisms that exploit information of any kind in the input may be innately specified.
2. Innate knowledge or constraints may specify, for instance, the number of syntactic categories or the relationships between them.

4. Utility of Information Sources' Access

- Distributional analysis can be conducted over electronically stored texts, represented purely as sequences of distinct words, and these are (at least for English) available to researchers in almost unlimited supply.

5. Relevant Distributional Approaches

1. Distributional Analysis in Linguistics.

5. Relevant Distributional Approaches

1. Distributional Analysis in Linguistics.
2. Neural Networks.

5. Relevant Distributional Approaches

1. Distributional Analysis in Linguistics.
2. Neural Networks.
3. Statistical Approaches to Language Learning.

5.1 Distributional Analysis in Linguistics

- Distributional linguists were interested in the discovery of language structure from corpora, purely from the point of view of providing a rigorous methodology for field linguistics.

5.1 Distributional Analysis in Linguistics

- Distributional linguists were interested in the discovery of language structure from corpora, purely from the point of view of providing a rigorous methodology for field linguistics.
- They conceived of language as an external cultural product, and <u>did not</u> consider it in a psychological or computational context.

5.1 Distributional Analysis in Linguistics

- Distributional linguists were interested in the discovery of language structure from corpora, purely from the point of view of providing a rigorous methodology for field linguistics.
- They conceived of language as an external cultural product, and <u>did not</u> consider it in a psychological or computational context.
- They were unable to test their methods except with very small samples of language.
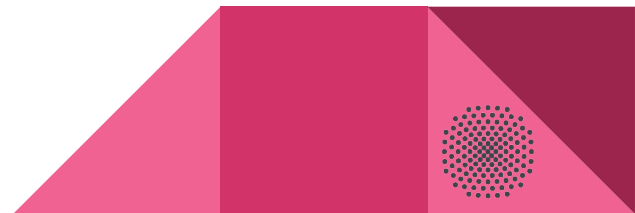
5.2 Neural Networks

- The most influential for learning the structure of a sequential material uses SRNs.

5.2 Neural Networks

- The most influential for learning the structure of a sequential material uses SRNs.

Simple Recurrent Networks

Assign similar hidden unit patterns to items which have the same syntactic category in a simple grammar.

5.2 Neural Networks

- Another approach for learning the linguistic categories of small artificial languages uses a <u>competitive network</u> in order to produce a topographic mapping between the distribution of contexts in which an item occurs and a 2-dimensional space.
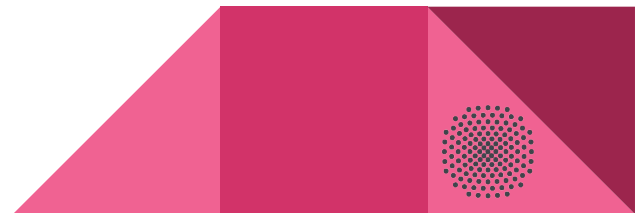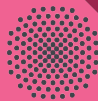
5.2 Neural Networks

- Another approach for learning the linguistic categories of small artificial languages uses a <u>competitive network</u> in order to produce a topographic mapping between the distribution of contexts in which an item occurs and a 2-dimensional space.

Results

The results show that items with the <u>same linguistic category</u> tend to lie in neighboring regions of the space.

5.2 Neural Networks

Limitations (SRNs and Competitive Networks)

- Scaling up still not being possible from very small artificial data sets in order to deal with real linguistic data.

5.2 Neural Networks

Limitations (SRNs and Competitive Networks)

- Scaling up still not being possible from very small artificial data sets in order to deal with real linguistic data.
- The linguistic categories can only be revealed using a subsequent cluster analysis.

5.3 Statistical Approaches to Language Learning

Main aim:

- Practical utility.

5.3 Statistical Approaches to Language Learning

Main aim:

● Practical utility.

Problem:

● It has not demonstrated utility of distributional information concerning syntactic categories for a very large and rich corpora.

6. New Distributional Approaches

1. Measuring the Distribution of Each Word.

6. New Distributional Approaches

1. Measuring the Distribution of Each Word.
2. Comparing the Distributions of Pairs of Words.

6. New Distributional Approaches

1. Measuring the Distribution of Each Word.
2. Comparing the Distributions of Pairs of Words.
3. Grouping Together Words with Similar Distributions.

6.1 Measuring the Distribution of Each Word

Context for a word

6.1 Measuring the Distribution of Each Word

Context for a word

- It may be defined simply in terms of the distribution of words which occur near the target word.

6.1 Measuring the Distribution of Each Word

Context for a word

- It may be defined simply in terms of the distribution of words which occur near the target word.

Measurements' Records

- A record of such statistics can be viewed as a <u>contingency table</u>.
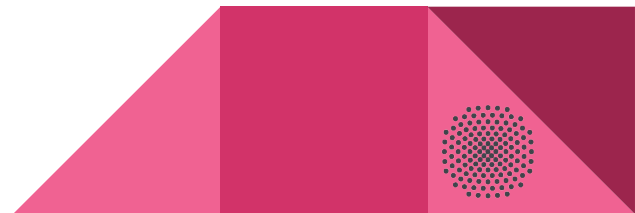
6.1 Measuring the Distribution of Each Word

Example:

6.1 Measuring the Distribution of Each Word

Example:

- **Input:** *The cow jumped over the moon.*
- **Target:** *jumped.*

6.1 Measuring the Distribution of Each Word

Example:

- **Input:** *The cow jumped over the moon.*
- **Target:** *jumped.*

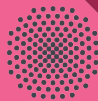**Indexed cells which would be incremented in the contingency table:**

6.1 Measuring the Distribution of Each Word

Example:

- **Input:** *The cow <u>jumped </u>over the moon.*
- **Target:** *jumped*.

Indexed cells which would be incremented in the contingency table:

(jumped, the), (jumped, cow), (jumped, over), (jumped, the), (jumped, moon).

6.2 Comparing the Distributions of Pairs of Words

- The more similar the <u>words' distributions</u>, the more likely that they are members of the same category.

6.3 Grouping Together Words with Similar Distributions

**Goal due to the syntactic categories' boundaries:**

6.3 Grouping Together Words with Similar Distributions

Goal due to the syntactic categories' boundaries:

- Assigning words to discrete categories.

6.3 Grouping Together Words with Similar Distributions

Goal due to the syntactic categories' boundaries:

- Assigning words to discrete categories.

Requirements for it:

6.3 Grouping Together Words with Similar Distributions

Goal due to the syntactic categories' boundaries:

- Assigning words to discrete categories.

Requirements for it:

- Non-hierarchical classification over the similarity space.
- Identifying <u>clusters</u> of similarly distributed target words.
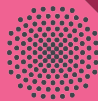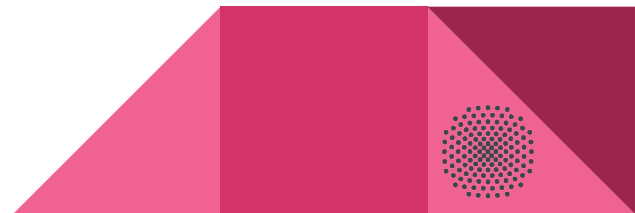
7. Accuracy and Completeness

Accuracy:

7. Accuracy and Completeness

Accuracy:

- Proportion of pairs of items that are grouped together in the derived groups which are also grouped together in the benchmark groups.

7. Accuracy and Completeness

Accuracy:

- Proportion of pairs of items that are grouped together in the derived groups which are also grouped together in the benchmark groups
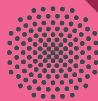
Completeness:

7. Accuracy and Completeness

Accuracy:

- Proportion of pairs of items that are grouped together in the derived groups which are also grouped together in the benchmark groups

Completeness:

- Proportion of pairs of items which are grouped by the benchmark that are also grouped together in the derived groupings.
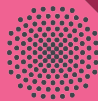
7. Accuracy and Completeness

Equations:

7. Accuracy and Completeness

Equations:

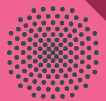$$Accuracy = \frac{hits}{hits + falseAlarms}$$

7. Accuracy and Completeness

Equations:

$$Accuracy = \frac{hits}{hits + falseAlarms}$$

$$Completeness = \frac{hits}{hits + misses}$$

8. Conclusion

- A model of how children may use distributional information in acquiring syntactic categories has been shown.

8. Conclusion

- A model of how children may use distributional information in acquiring syntactic categories has been shown.
- Distributional information is a potentially powerful cue for learning syntactic categories.

8. Conclusion

- A model of how children may use distributional information in acquiring syntactic categories has been shown.
- Distributional information is a potentially powerful cue for learning syntactic categories.
- The use of distributional methods is often associated with <u>empiricist approaches</u> to language acquisition.
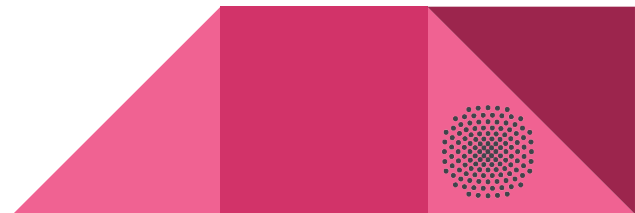
Thank you!

Distributional Information:
A Powerful Cue for Acquiring Syntactic Categories

Get the slides from this presentation on **hyperurl.co/distinfo**

Levindo Gabriel Taschetto Neto
22.11.2017

Probabilistic models of language and cognition

Universität Stuttgart