

Een Computatoneel Model Voor Taalacquisitie

Andreas van Cranenburgh

Arjan Nusselder

Nadya Peek

Carsten van Weelden

Universiteit van Amsterdam

June 29, 2007

Introductie

Waarom zijn we geïnteresseerd in taalacquisitie?

- Fundamentele eigenschap van mensen.
- AI probleem.
- Chomsky's innate grammar is niet aannemelijk.
- Natuurlijke taal voor Turing test.
- Een praktische aanpak.

Overzicht

- Achtergrond
- Hypothese
- Taakverdeling
- Implementatie
- Resultaten
- Conclusie

Achtergrond

Generatieve & constructieve grammatica

- **Chomsky**: generatieve grammatica. Taalacquisitie met behulp van een aangeboren grammatica. Magerheid van stimulus argument.
- **Tomasello**: constructieve grammatica. Magerheid van stimulus argument gaat niet meer op.

Motivatie

Beide partijen hebben nog geen empirisch bewijs

Experiment: computationeel **model van constructieve grammatica**

Verwerking:

- Een corpus om analogiën mee te maken

Verwerving:

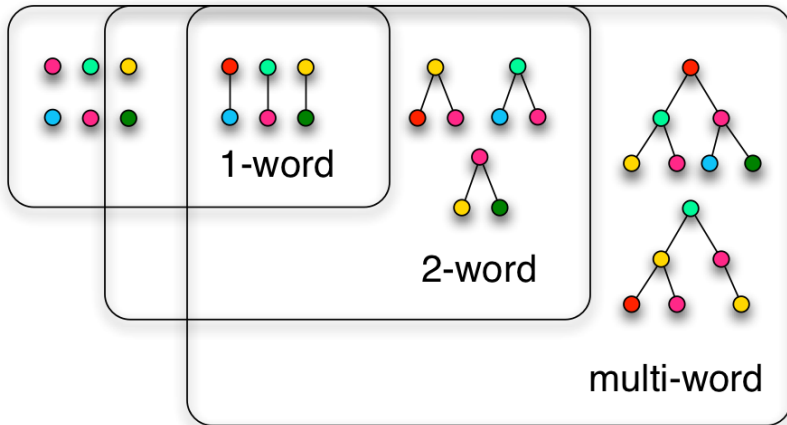
- Het verkrijgen van het eerste corpus

Kernvraag

Hoe verkrijgen kinderen hun **eerste corpus** om analogiën mee te kunnen maken?

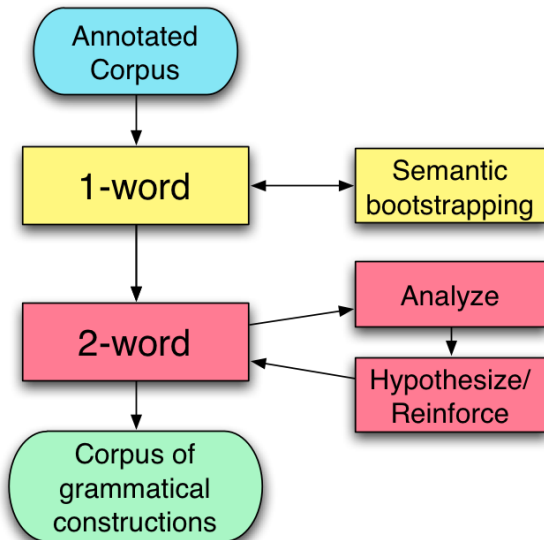
Hypothese

Zij doen dit in opeenvolgende fases:



Hypothese

Het is mogelijk een **computationeel model** te bouwen, dat door middel van deze stappen **taal verwerft**.



Taakverdeling

Grove taakverdeling:

- Two word stadium specificatie
- One word stadium implementeren
- Two word stadium implementeren
- Corpus annoteren
- Experimenten draaien

Samenwerking:

- wiki voor gezamenlijke geklad
- code repository op SVN server
- Trac voor planning, bugtracking en taakverdeling

Semantische representatie

- Uit prelinguïstische fase
- Sterk gekarikaturiseerd
- Voorbeeld situatie:

DESC: mother points to block

ADULT UTTERANCE: "yes !that's a !block"

FRAME: action

 ID: point

 ABSTR: focus

 FRAME: object

 ID: block

 ABSTR: object:toy

 PROP: shape = square

 FRAME: who

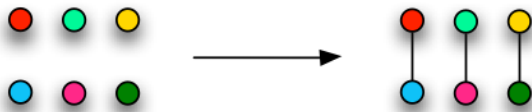
 ID: mother

 ABSTR: object:human

 PROP: family = yes

 PROP: gender = female

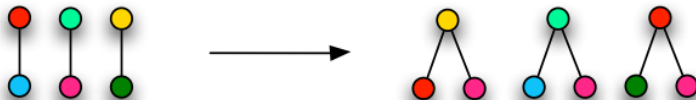
Aanpak I



One word Leren van losse woorden en bijbehorende betekenissen.

- *Derive Meanings:*
 - Genereer subframes uit situaties.
 - Extraheer properties.
- *Scoringsmethode:*
 - Associeer (sub)frames met woorden dmv statistiek.
 - Corrigeer voor veelvoorkomende frames / woorden.

Aanpak II



Two word

- Maakt **linguïstische abstracties** op basis van frames.
- Construeren van woord paren op basis van abstracties.
- Vormt corpus van **linguïstische constructies**.
- Sleutelen aan scoringsmethode geeft gemengde resultaten.

Resultaten I



Nieuwe Betekenissen

- Linguïstische abstracties kunnen worden gecombineerd.

Demo

- *Simpel corpus*
- Voorbeeld van linguïstische abstracties die worden gecombineerd.

Resultaten II

Nieuwe Woordcombinaties

- Tests met woordgeneratie op geleerde situaties levert nieuwe woordcombinaties op.
- Ruim de helft (58%) van de uitspraken kwamen niet voor in het corpus.
- Niet alle uitspraken lijken even betekenisvol.

Resultaten III

Voorbeelden van gegenereerde woord combinaties:

Situation: 24

1 : ball give Score = 209

2 : give ball Score = 129

Situation: 157

1 : hold you Score = 29

2 : hold still Score = 29

Conclusie



- De hypothese dat nieuwe woord-combinaties geleerd kunnen worden is niet weerlegd.
- Het aanleren van taal middels “performance based” methoden lijkt in principe mogelijk.