

Übung zur Einführung in die Computerlinguistik

Wiebke Petersen

WiSe 04/05

Turing-Test

1950, Alan Turing, *imitation game*:

Testkandidat muß per Fernschreiber zwischen männlichen und weiblichen Gesprächspartnern unterscheiden.

"Kommunikationspartner" wird durch eine Maschine ersetzt, die entweder einen männlichen oder weiblichen Gesprächspartner simuliert.

Die Maschine besteht den "Turing-Test", wenn sie einen menschlichen Gesprächspartner so gut simuliert, daß der Testkandidat die Maschine ebenso oft richtig bzw. falsch klassifiziert wie den menschlichen.

1.5 ELIZA – Ein Beispieldialog

- ELIZA (Weizenbaum, 1966) ist ein Dialogsystem, das einen Psychologen simuliert und mit natürlichsprachlichen Äußerungen auf natürlichsprachliche Eingaben reagiert
- Ein Beispieldialog:
 - User: Men are all alike.
 - ELIZA: IN WHAT WAY
 - User: They're always bugging us about something or other.
 - ELIZA: CAN YOU THINK OF A SPECIFIC EXAMPLE
 - User: Well, my boyfriend made me come here.
 - ELIZA: YOUR BOYFRIEND MADE YOU COME HERE
 - User: He says I'm depressed much of the time.
 - ELIZA: I AM SORRY TO HEAR YOU ARE DEPRESSED

1.5 ELIZA – Die Funktionsweise

- kein wirkliches „Verstehen“, sondern **Mustererkennung**
- Konstruktion von Antwortphrasen und Antwortsätzen abhängig vom gefundenem Muster
- Beispiele:
 - alle Vorkommen von *I'm* werden zunächst durch *YOU ARE* substituiert und alle Vorkommen von *my* durch *YOUR*
 - *YOU ARE depressed* ⇒ *I AM SORRY TO HEAR THAT YOU ARE DEPRESSED*
 - *ALL* ⇒ *IN WHAT WAY*
 - *ALWAYS* ⇒ *CAN YOU THINK OF A SPECIFIC EXAMPLE*

1.5 ELIZA – Problem: Morphologie

- reichere Morphologie des Deutschen gegenüber dem Englischen ⇒ „deutschsprachiges ELIZA“ wesentlich aufwendiger
 - Beispiel [*my* ⇒ *YOUR*]:
 - mein ⇒ DEIN
 - meine ⇒ DEINE
 - meinen ⇒ DEINEN
 - ...
- ⇒ Morphologische Analysekomponente selbst für ein primitives Programm wie ELIZA unerlässlich

Kompetenz und Performanz

Sprachliche Kompetenz:

- die endliche strukturierte Wissensbasis, die es den Sprechern einer Sprache ermöglicht, die wohlgeformten Äußerungen der Sprache zu generieren und zu interpretieren.

Sprachliche Performanz:

- die Generierung oder Interpretation realer Äußerungen, bzw. die Gesamtheit der Prozesse, die beteiligt sind, wenn der Mensch auf der Basis der sprachlichen Kompetenz reale Äußerungen generiert und interpretiert.

Kompetenzmodell

Ein Kompetenzmodell sollte beinhalten:

Regeln, Prinzipien, Beschränkungen auf jeder Beschreibungsebene, die in ihrem Zusammenwirken genau die wohlgeformten Sätze der Sprache charakterisieren.

Es bietet für jede Sprache eine formalisierte endliche Definition einer unendlichen Menge von Paaren <Satz, Bedeutung>.

(Dazu gehören: Grammatik, Lexikon, morphologische Regeln, semantische Regeln.)

Performanzmodell

Ein Performanzmodell sollte erklären:

warum viele ungrammatische Sätze erzeugt werden

- z.B. Sprechfehler, Grammatikfehler

warum viele ungrammatische Sätze verstanden werden

- z.B. in der der Kommunikation mit Kindern oder Ausländern

warum viele grammatische Sätze nicht erzeugt werden

- z.B. durch Präferenzen in der Generierung

warum viele grammatische Sätze nicht verstanden werden

- z.B. Holzwegsätze

wie die Verarbeitung zeitlich strukturiert ist

- z.B. Effizienz, Abfolge der Verarbeitungsschritte

welchen Aufwand die Verarbeitungsschritte erfordern

- z.B. Abhängigkeiten von anderen kognitiven Belastungen

Beispiele

Grammatikfehler und Sprechfehler:

- Das Verfassen der Kinderbücher und der Reiseberichte haben dem Autor viel Ruhm eingebracht.
- Die Poxen zum Backen...

Holzwegsätze:

- The canoe floated down the river sank.
- Er bezichtigte den Vater des Schreibens unkundiger Kinder.
- Peter beschuldigte sie der Geheimniskrämerei ähnlichen Verhaltens.

Ambiguität

phonetische Ambiguität (Homophone)

Miene - Mine

orthographische Ambiguität (Homographen)

übersetzen - übersetzen

lexikalische Ambiguität (Homonyme)

Ball - Ball

morphologische Ambiguität

Staubecken - Staubecken

Hauptpostsekretär

Lexikalische Ambiguität

Gewisse Lesarten sind weniger stark präferiert:

Auf dem Tisch lag ein Heft.

Auf der Werkbank lag ein Heft.

Die Präferenz für eine Lesart kann durch den Kontext beeinflusst werden:

Der Mittelstürmer eröffnete den Ball.

versus

Der Präsident eröffnete den Ball.

Der Gärtner sprengte das Schloß.

versus

Der Einbrecher sprengte das Schloß.

The astronomer married a star.

versus

The movie director married a star.

Ambiguität II

syntaktische Ambiguität

Peter fuhr seinen Freund sturzbetrunken nach Hause.

Visiting relatives can be boring.

Ich traf den Sohn des Nachbarn mit dem Gewehr.

kompositionell-semantische Ambiguität

Die zwei Mitarbeiter müssen vier Sprachen beherrschen.

pragmatische Ambiguität

Könnten Sie die Aufgabe lösen.

Ambiguität (Beispiel)

„Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.“

Ambiguität (Beispiel)

„Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.“

Der Satz weist lexikalische (L), syntaktische (S) und anaphorische (A) Ambiguitäten auf, die uns nicht auffallen.

Ambiguität (Beispiel)

„Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.“

Der Satz weist lexikalische (L), syntaktische (S) und anaphorische (A) Ambiguitäten auf, die uns nicht auffallen.

Wieviele Lesarten besitzt dieser Satz?

258.048

Ambiguität (Beispiel)

„Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.“

Das berechnet sich so:

- L** *Früher* kann sowohl eigenständiges Adverb als auch Komparativ von *früh* sein (2);
- L** die Verbform *stellten* ist ambig zwischen Präteritum und Konjunktiv (2);
- S** die Nominalphrase *die Frauen* kann sowohl Subjekt als auch Objekt des Satzes sein (2);
- S** *am Wochenende* kann die Insel, die Frauen oder das Verb modifizieren (3);
- S** *mit Blumenmotiven* kann sich auf die Kopftücher beziehen, ein Instrument der Herstellung sein oder ein Adjunkt im Sinne von *gemeinsam mit Blumenmotiven* (3);
- L** *her* hat auch eine direktionale Bedeutung (2);

Ambiguität (Beispiel)

„Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.“

Und weiter:

- S** der Relativsatz könnte jede der vier Nominalphrasen im Plural modifizieren (4);
- S** sowohl *die* als auch *ihre Männer* kann Subjekt des Relativsatzes sein (2);
- A** das Possessivpronomen *ihre* kann auf jede der Nominalphrasen referieren (4);
- L** *Montagen* hat eine zweite Lesart als Nominalisierung von *montieren* (2);
- S** *die Hauptinsel* kann im Genitiv zu der vorangegangenen NP gehören oder im Dativ die Käuferin bezeichnen (2);
- S** die drei Präpositionalphrasen des Relativsatzes können sich in insgesamt sieben Kombinationen mit den jeweils vorhergehenden NPs oder mit dem Verb verbinden (7);
- L** *verkauften* zeigt wieder die Ambiguität zwischen Präteritum und Konjunktiv auf (2).

Ambiguität (Beispiel)

„Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.“

Durch Multiplikation ergibt sich die Gesamtambiguität:

$$2 \times 2 \times 2 \times 3 \times 3 \times 2 \times 4 \times 2 \times 4 \times 2 \times 2 \times 7 \times 2 = \underline{258.048}$$

Referate (1)

Korrekturprogramme: Rechtschreibkorrektur, Grammatikkorrektur
Korrekturvorschläge, Verbesserung von Texterfassung mittels OCR.

Computergestützte Lexikographie: Hilfe bei der Erstellung und Pflege von Lexika; Akquisition lexikalischer Information, Repräsentation lexikalischer Information, Bereitstellung der lexikalischen Information für Anwendungen.

Volltextsuche (Information Retrieval): Indexkonstruktion, Auswertung von Suchanfragen, Retrievalmodell

Textmining: Strukturierung großer Textkollektionen, Textklassifikation, Schlüsselwortextraktion, Aufbau einer Taxonomie

Textklassifikation: Erlernen von Klassenprofilen anhand von Trainingsdaten, Klassifikationsalgorithmus

Informationsextraktion: Identifizierung relevanter Information in Texten, Instantiierung von Templates

Textzusammenfassung: Reduktion / Verdichtung, Textproduktion

Referate (2)

Sprachsynthesysteme: Produktion gesprochener Sprache aus geschriebener Sprache, Computerarbeitsplatz für Blinde, telefonische Auskunftssysteme, Navigationsysteme

Spracherkennungssysteme: Diktiersysteme, telefonische Auskunftssysteme; Signalanalyse, Geräuschfilterung, Adaption an verschiedene Sprecher

Natürlichsprachliche Retrieval-Schnittstellen: z.B. natürlichsprachliche Anfragen an Bibliothekskataloge

Dialogsysteme: ELIZA, automatische Auskunftssysteme, natürlichsprachliche Benutzerschnittstellen

Sprachlehr- und -lernsysteme: Hilfe bei dem Erwerb von Fremdsprachen; Anpassung an das individuelle Arbeitstempo und den Kenntnisstand, ortsungebunden, zeitlich flexibel, objektiv, nicht ermüdend

Referate (3)

Elektronische Kommunikationshilfen: Wort- und Satzvervollständigung (SMS), Texttelephone, Textvereinfachungswerkzeuge, ...

Angewandte natürlichsprachliche Generierungs- und Auskunftssysteme: Wettervorhersagen, Gesundheitswesen, technische Dokumentationen, Computerspiele, ...

Maschinelle Übersetzung: Vollautomatische Übersetzung, Computergestützte Übersetzung

Übungsaufgabe

Internetrecherche zu computerlinguistischen
Instituten

- Wo befinden sich wichtige CL Institute?
- Wer leitet sie?
- Wo liegen ihre Forschungsschwerpunkte?

Erstellen sie eine kurze Powerpoint-Präsentation
zu einem von ihnen ausgewählten Institut.