

Determination der Informationsstruktur in einem Dialogsystem

Kepa J. Rodríguez

8. Juli 2003

1 Informationsstruktur-Partitionierung: Steedmans Theorie

Es wurde die Theorie von Steedman ([Ste00]) verwendet, weil sie eine explizite Formalisierung der Informationsstruktur-Determination anbietet.

Das Ziel der Theorie ist eine kompositionale Analyse der englischen Intonation mit Hilfe der Informationsstruktur.

Hauptmerkmale der Theorie sind:

- Bidimensionale IS-Partition in Thema/Rhema und Fokus/Background.
- Einfache Modellierbarkeit für die Sprachsynthese.

1.1 Die Informationsstruktur-Partition

Die zwei Dimensionen der Informationsstruktur-Partition sind:

1. Die Thema/Rhema-Partitionierung: Ergibt sich auf Satzebene. 'Aboutness'-Relation, in der das Rhema semantisch über das Thema prädiziert (semantically predicated).
2. Die Fokus/Hintergrund-Partitionierung: Kontrast zwischen Alternativen im Diskurskontext.

- (1) **Q:** I know that Marcel likes the man who wrote the musical.
But who does he ADMIRE?
- A:** (Marcel ADMIREs) ^{LH%} (the woman who DIRECTED the musical)
- $\underbrace{\text{background} \quad \text{L+H*}}_{\text{focus}} \quad \text{LH\%} \quad \underbrace{\text{background} \quad \text{H*}}_{\text{focus}} \quad \underbrace{\text{LL\%}}_{\text{background}}$
- $\underbrace{\hspace{15em}}_{\text{theme}} \quad \underbrace{\hspace{15em}}_{\text{rheme}}$

Wir können die Semantik der Informationsstruktur-Partitionierung folgendermaßen definieren:

- Rhema: präsupponiert eine Rhema-Alternativmenge (ρ -AS). Fokus-Rhema beschränkt die ρ -AS auf eine ein-elementige Menge.
- Thema: präsupponiert eine Thema-Alternativmenge (θ -AS). Ohne Fokus-Thema hat die θ -AS nur ein Element. Ansonsten beschränkt Fokus-Thema die θ -AS auf eine ein-elementige Menge.

1.2 Informationsstruktur und Intonation

Laut Steedman stehen die IS und die intonationale Struktur im Englischen in einer homomorphen Relation:

- Thema/Rhema-Partition determiniert das allgemeine Intonationsmuster, d.h., welche Akzente und Grenztöne innerhalb jeder Partition benutzt werden können, sowie die Positionen, wo Grenztöne auftreten können.
Das Intonationsmuster für English ist:
 - Rhema-Kontur: H* LL%
 - Thema-Kontur: L+H* LH%
- Fokus/Hintergrund-Partition determiniert die Positionen der Pitch-Akzente. Wörter, die zum Hintergrund gehören, werden nicht akzentuiert.

2 Dialogmanagement: ein paar Grundlagen

2.1 Information States Theory

Die wichtigsten Elemente der Theorie ([LT00]) sind:

- **Information States:**([Coo98])
 - Attribut-Wert-Matrix (siehe Abb. 1).
 - Einziger Zustand für ganzen Dialog.
 - Kontext, Kontributionen, mentale Zustände, usw. werden repräsentiert.
 - Nach jeder Kontribution wird der Information State updated.
- **Update Regeln:**
 - Organisiert als Produktionssystem.
 - 2 Felder: Vorbedingungen und Effekte.
- **Update Strategie:**
 - Entscheidet, welche Regel aus einer Menge von anwendbaren Regeln angewendet wird.

Als Beispiel in Abb. 1 die Repräsentation des Information State in GoDiS (Gothenburg Dialogue System [LAJQ02]).

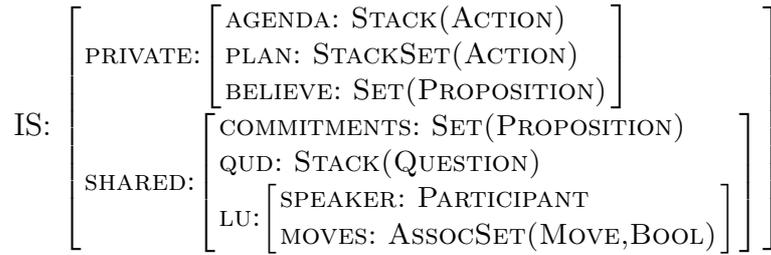


Abbildung 1: Information State in GoDiS [LAJQ02]

2.2 Question Under Discussion (QUD)

QUD ist eine Menge, die die zur Diskussion stehenden Fragen im aktuellen Diskurs spezifiziert. Sie sind nach konversationaler Präzedenz geordnet. Wir können das als Stapel repräsentieren. Wenn eine Frage q maximal in QUD ist (ganz oben auf dem Stapel), ist es erlaubt, q -spezifische Information durch eine (optional) kurze Antwort zu geben ([Gin96]).

- Wenn ein Dialog-Partizipant eine Frage stellt, wird sie auf den QUD-Stapel gelegt.
- Wenn eine Frage beantwortet wird, wird sie vom QUD-Stapel entfernt.

3 IS-Determination Regeln

Im Rahmen des Projektes SIRIDUS (Specification, Interaction and Reconfiguration in Dialogue Understanding Systems) wurden folgende Regeln entworfen ([KKEG⁺02]), die bestimmte Informationstruktur-Partitionierungen erklären können:

- **QudTR:** QUD-basierte Thema-Rhema Determination.
- **ComFB:** Commitment-basierte Fokus-Hintergrund Determination.
- **DomFB:** Domänen-basierte Fokus-Hintergrund Determination.

In Abb. 2 finden wir den Hauptalgorithmus, der die Reihenfolge der Anwendung der Regeln entscheidet.

3.1 QUD-basierte Thema-Rhema Determination (QudTR)

1. wenn
 - (a) Frage q liegt ganz oben auf dem QUD-Stapel
 - (b) eine Aussage u mit semantischem Inhalt c wird geäußert.
 - (c) q ist eine λ -Abstraction über c

2. dann

- (a) der Teil von c , der mit q korrespondiert, gehört zum Thema von u
- (b) der Rest ist der informative Teil und gehört zum Rhema von u

Beispiel:

- (2) **U:** Wieviel kostet der Flug?
 S: $\underbrace{\text{Der Flug kostet}}_{\text{Thema}} \underbrace{500 \text{ Euro}}_{\text{Rhema}}.$
 QUD vor der Antwort: $?\lambda x. \text{Preis}(x)$
 Propositionaler Inhalt der Antwort: $\text{Preis}(500_Euro)$
 Thema-Teil der Antwort: $\lambda x. \text{Preis}(x)$
 Rhema-Teil der Antwort: 500_Euro

3.2 Commitment-basierte Fokus-Hintergrund Determination (ComFB)

Commitments: Propositionen, auf die sich die Dialog-Partizipanten festgelegt haben.

1. wenn

- (a) die Aussage u hat den Inhalt c
- (b) in SHARED COMMITMENTS haben wir die Proposition p
- (c) p und c sind parallel

2. dann

- (a) jedes Element in einer Informationseinheit von c , das eine Alternative in der parallelen Informationseinheit p hat, wird mit Fokus markiert.

Beispiel:

- (3) **U1:** Wieviel kostet erste Klasse?
 S1: Erste Klasse kostet 1000 Euro
 In SHARED COMMITMENTS: $\text{Price}(1000, Euro), \text{Class}(Business)$
 U2: Und zweite Klasse?
 QUD: $?\lambda x. \lambda y. \text{Preis}(x, y)$
 Antwort: $\text{Price}(500, Euro), \text{Class}(Economy)$
 S2: $\underbrace{\underbrace{\text{Zweite Klasse kostet}}_{\text{Focus}} \underbrace{500}_{\text{Background}}}_{\text{Thema}} \underbrace{\underbrace{\text{Euro}}_{\text{Focus}}}_{\text{Background}}.$

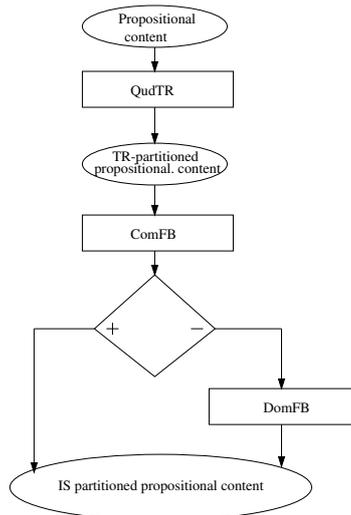


Abbildung 2: Architektur der Informationsstruktur-Determination

3.3 Domänen-basierte Fokus-Hintergrund Determination (DomFB)

1. wenn
 - (a) die Aussage u hat den Inhalt c
 - (b) ein Element $e \in c$ hat eine Alternativ e' in der Domäne.
2. dann
 - (a) e wird mit Fokus markiert.

Beispiel:

- (4) **U:** Wieviel kostet der Flug?
(S: $\underbrace{\text{Der Flug kostet}}_{\text{Thema}} \underbrace{500 \text{ Euro.}}_{\text{Rhema}}$
 Antwort: Preis(500,Euro)
 In Domäne: Preis(300,Euro), Preis(1000,Euro)...
- S:** $\underbrace{\text{Der Flug kostet}}_{\text{Thema}} \underbrace{500}_{\text{Focus}} \underbrace{\text{Euro}}_{\text{Background}} \text{ .}$
 $\underbrace{\hspace{10em}}_{\text{Rhema}}$

4 Anwendung in Dialogsystemen

Informationsstruktur-Determination hat mehrere Anwendungen in Dialogsystemen. Unter anderem finden wir:

- Verbesserung der Intonation im System-Output.
- Entscheidung über Konstituentenreihenfolge im System-Output.
- Resolution von anaphorischen Referenzen.
- ...

4.1 Output von verschiedenen TTS-Systemen

Die meisten text-to-speech (TTS) Systeme nutzen nur syntaktische Information und den Satzmodus, um die Intonation zu modellieren. Die Wirkung davon ist, dass wenn wir nur einen isolierten Satz hören, er oft sehr gut klingt. Die Intonation ist aber sehr künstlich, wenn wir den Satz im Kontext betrachten.

Betrachten wir die Ausgaben mehrerer freier und kommerzieller TTS-Systeme ([KKEG⁺02]):

- When do you WANT to leave? Festival und ViaVoice
- When DO you want to leave? Lucent's Articulator
- When do you want TO leave? AT&T's TTS

4.2 Beispiel: Anwendung in SIRIDUS

Das Projekt SIRIDUS¹ (Specification, Interaction and Reconfiguration in Dialogue Understanding Systems) ist ein Nachfolgerprojekt von Trindi².

Die Projektpartner waren:

- Universitäten: Göteborg, Saarland, Sevilla.
- Wirtschaft: Linguamatics (Schottland), Telefonica I+D (spanischer Staat).

Die Determination der Informationsstruktur wurde in der lokalen Forschungsgruppe der Universität des Saarlandes³ in Zusammenarbeit mit Göteborg entwickelt.

Merkmale des Teilprojekts sind:

- Verwendetes Dialogsystem: GoDiS (Gothenburg Dialogue System [LE02]).
- Sprachen: Deutsch und Englisch.
- Verwendeter phonologischer Repräsentationsformalismus: ToBI (Englisch) und GToBI (Deutsch).

¹<http://www.ling.gu.se/projekt/siridus>

²<http://www.ling.gu.se/projekt/trindi>

³<http://www.coli.uni-sb.de/cl/projects/siridus>

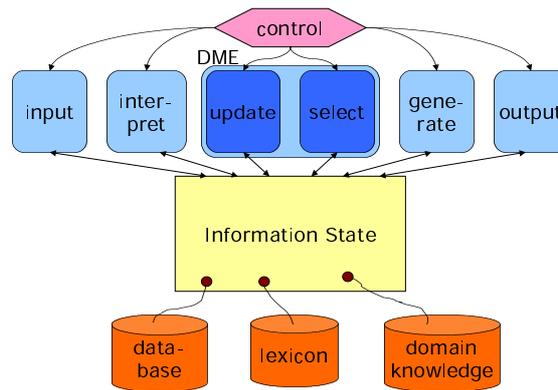


Abbildung 3: Architektur von GoDiS ([LAJQ02])

- TTS-Systeme: Flexiblere Systeme, wo die Intonation von draußen manipulierbar ist. Sie akzeptieren ToBI Notation.
 - Festival⁴ (CSTR-Edinburgh): Für Englisch. Mark-up: APML⁵.
 - Mary⁶ (DFKI): Für Deutsch. Mark-up: MaryXML.

Die Abbildung 4 zeigt, wie die TTS-Systeme mit der Informationsstruktur-Partitionierung integriert wurden.

4.3 GoDiS: Gothenburg Dialogue System

Information State basierte Dialog-System (Abb. 3).

Domäne: Reiseagentur, intelligentes Haus, Videorecorder

Sprachen: Englisch, Deutsch, Schwedisch.

4.4 ToBI und GToBI

ToBI ([BH94]) ist ein Notationssystem für die Transkription der Intonation.

Die Bestandteile des Systems sind:

- Pitch-Akzente: Repräsentiert durch '*’.
- Grenztöne
 - Grenze einer intermediären Phrase: Repräsentiert durch ‘-’.
 - Grenze einer intonationalen Phrase: Repräsentiert durch ‘%’.

Die verschiedenen Pitch-Akzente und Töne werden durch Kombination von 2 Grundtönen geschaffen: L (low) and H (high).

Für die deutsche Version wurde GToBI ([GBB]) -ein ToBI für Deutsch- verwendet.

⁴<http://www.cstr.ed.ac.uk/projects/festival>

⁵Entwickelt von Robert Clark <http://www.cstr.ed.ac.uk/~robert>

⁶<http://mary.dfki.de>

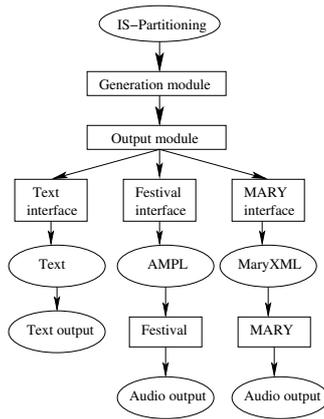
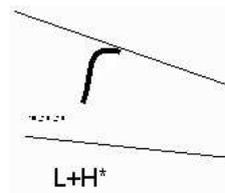
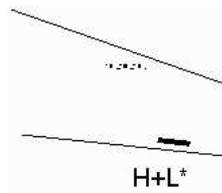


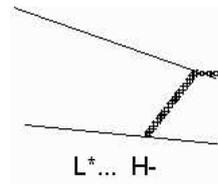
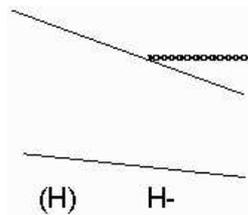
Abbildung 4: Schnittstellen zwischen GoDIS und TTS-Systemen ([KKKRE03])

Die ausgewählten GToBI-Akzente sind:

- Verwendete GToBI Akzente:



- Verwendete GToBI Grenztöne



In der Tabelle 1 können wir die Korrespondenz zwischen GToBI-Akzenten und -Grenztönen mit informationsstrukturellen Kategorien sehen.

IS-Partitionierung	GToBI
Fokus-Thema	L+H*
Fokus-Rhema	H+L*
Unmarkierte Thema-Grenze (vor Rhema)	-
Markierte Thema-Grenze (vor Rhema)	H- (break index 3)
Rhema-Grenze (vor Thema)	-

Tabelle 1: Korrespondenz Informationsstruktur/GToBI

4.5 Evaluation

4.5.1 Perzeptionsexperiment

Die Evaluation der Implementierung ([KKERK03]) wurde durch ein Perzeptionsexperiment gemacht.

Die TeilnehmerInnen sollten Dialogfragmente lesen und einen Satz hören. Sie sollten bewerten, wie gut die Intonation des Satzes im gegebenen Kontext ist. Die Skala verlief von 1 (sehr schlecht) bis 5 (ausgezeichnet).

Gleiche Beispiele wurden mit und ohne Informationstruktur-basierte Manipulation produziert.

Ein Problem war, dass viele Probanden die Neigung hatten, alle Beispiele sehr gut oder sehr schlecht zu bewerten. Das wurde durch die Berechnung des Durchschnitts der Differenzen zwischen der Punktzahl von manipulierten und nicht manipulierten Outputs gelöst.

Wie wir in Tabelle 2 sehen können, bevorzugten die Probanden die manipulierten Outputs.

	Th vor Rh	Rh vor Th
+Focus-Thema	+0.29	+0.28
-Focus-Thema	+0.3	+0.26

Tabelle 2: Ergebnisse des Perzeptionsexperiments

Literatur

- [BH94] Beckman and Julia Hirschberg. The ToBI annotation conventions. Technical report, Ohio State University, USA, 1994.
- [Coo98] Robin Cooper. Information States, Attitudes and Dialogue. In R. Cooper and T. Gamkrelidze, editors, *Proceedings of the Second Tbilisi Symposium on Language, Logic and Computation*, Tbilisi, Georgia, 1998.
- [GBB] Martine Grice, Stefan Baumann, and Ralf Benzmüller. German intonation in autosegmental-metrical phonology. In Jun Sun-Ah, editor, *Prosodic Typology*. Oxford University Press.
- [Gin96] Jonathan Ginzburg. Interrogatives: Questions, facts and dialogue. In *The Handbook of Contemporary Semantic Theorie*. Blackwell, Oxford, UK, 1996.
- [KKEG⁺02] Ivana Kruijff-Korbayová, Stina Ericsson, Carlos García, Rebecca Jonson, Elena Karagjosova, Pilar Manchón, Kepa J. Rodríguez, and José Quesada. Improving system output using the information state. Deliverable D5.1, SIRIDUS, December 2002.

- [KKERK03] Ivana Kruijff-Korbayová, Stina Ericsson, Kepa J. Rodríguez, and Elena Karagjosova. Producing contextually appropriate intonation in an information-state based dialogue system. In *Proceedings of the 10th Conference of the European Chapter of the ACL*, Budapest, Hungarn, April 2003. forthcoming.
- [KKKRE03] Ivana Kruijff-Korbayová, Elena Karagjosova, Kepa J. Rodríguez, and Stina Ericsson. A dialogue system with contextually appropriate spoken output information. In *Proceedings EACL'03*, Budapest, Hungarn, April 2003.
- [LAJQ02] Staffan Larsson, Gabriel Amores, Rebecca Jonsson, and José Quesada. Siridus system architecture and interface report (enhanced version). Deliverable D6.3, SIRIDUS, 2002. URL: <http://www.ling.gu.se/projekt/siridus>.
- [LE02] Staffan Larsson and Stina Ericsson. Issue Based Dialogue Management in a Multi-Domain, Multi-Language Dialogue System. In R. Smith, editor, *Demonstration Abstracts, ACL-02*, 2002.
- [LT00] Staffan Larsson and David Traum. Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. *Natural Language Engineering*, (1), 2000.
- [Ste00] Mark Steedman. Information Structure and the Syntax-Phonology Interface. *Linguistic Inquiry*, (31):649–689, 2000.