

Automatisches Dolmetschen mit Mehr-Ebenen-Charts

Jan W. Amtrup

Computing Research Laboratory
New Mexico State University
email: jamtrup@crl.nmsu.edu

Das menschliche Sprachverstehen arbeitet inkrementell. Wir beginnen bereits, Teile der Äußerung eines Gesprächspartners zu verstehen, bevor dieser ausgereedet hat. Folglich muß ein System, von dem ähnlich natürliche Leistungen erwartet werden (etwa beim automatischen Simultandolmetschen), ebenfalls die Eigenschaft besitzen, Eingabe in kleinen Abschnitten verarbeiten zu können. Das in diesem Artikel vorgestellte System MILC (*Machine Interpreting with Layered Charts*) repräsentiert ein durchgehend inkrementelles Programm zur automatischen Übersetzung natürlicher, spontan gesprochener Sprache. Es basiert neben der Inkrementalität auf zwei weiteren Architekturprinzipien, nämlich der Integriertheit der Wissensrepräsentation in einer Mehr-Ebenen-Chart, und der Uniformität der Verarbeitung mit Hilfe eines linguistischen Formalismus, der auf der Unifikation komplexer, getypter Merkmalstrukturen beruht.

MILC wurde anhand von fünf Terminabsprache-Dialogen getestet, und erzielt eine Erfolgsrate von 60% approximativ korrekten Übersetzungen. Dabei operiert es lediglich um 17% weniger effizient als eine vergleichbare nicht-inkrementelle Variante, und arbeitet in sechsfacher Echtzeit.

1 Einführung

Das menschliche Sprachverstehen arbeitet inhärent inkrementell. Damit ist gemeint, daß wir Teile einer Äußerung bereits zu verstehen beginnen — und gar auf sie reagieren — bevor ein Gesprächspartner ausgereedet hat. Diese Eigenschaft ermöglicht es uns, einem kontinuierlichen Strom von Sprachsignalen zu folgen. Konferenzdolmetscher gehen sogar einen Schritt weiter und produzieren den gehörten Inhalt gleichzeitig in einer anderen Sprache.

Psycholinguistische Forschungen haben etliche Bereiche und Prozesse isoliert, deren Inkrementalität wesentlich zur Leistungsfähigkeit des Sprachverstehens des Menschen beiträgt. So geht z.B. die Kohorten-Theorie [MWT80] davon aus, daß die Erkennung von Wörtern schon mit einem kleinen Ausschnitt des akustischen Signals (etwa 150 Millisekunden) beginnt. Basierend auf einem solchen kurzen Wortanfang wird eine Menge von potentiell gesprochenen Wörtern hypothetisiert, deren Umfang durch den weitergehenden Input reduziert wird. Bleibt lediglich ein Wort in der Menge (der sog. *Kohorte*) übrig, so wird angenommen, dies sei das tatsächlich gesprochene Wort. Im Extremfall führt das dazu, daß ein Wort erkannt wird, bevor es komplett ausgesprochen ist. Auf der höheren Ebene der grammatischen Verarbeitung spielen ebenfalls Faktoren eine Rolle, die nur inkrementell Einfluß auf das Verstehen einer Äußerung haben können. So zeigt etwa Niv [Niv93], daß neue Nominalphrasen, die bisher nicht im Kontext in Erscheinung getreten sind, bevorzugt als Objekt eines Satzes verstanden werden, nicht jedoch als dessen Subjekt.

Es scheint somit bereits aus Gründen einer Verarbeitungsadäquatheit sehr wünschenswert, das Paradigma der Inkrementalität ebenfalls in artifizielle System zur Verarbeitung natürlicher Sprache zu integrieren. Aber selbst wenn solchermaßen prinzipielle Gründe wenig Gewicht hätten, so eröffnet ein inkrementelles Vorgehen neue Wege und eine höhere Qualität natürlichsprachlicher Systeme:

- Wenn mehrere Module eines komplexen sprachverarbeitenden Systems auf kleinen Ausschnitten arbeiten können, so ergibt sich automatisch die Möglichkeit, eine *intermodulare Parallelität* auszunutzen und verschiedene Aspekte einer Äußerung gleichzeitig zu bearbeiten.
- Ein inkrementelles Vorgehen in einem modularen System birgt ebenfalls die Möglichkeit der gegenseitigen Beeinflussung von Teilen einer Anwendung. Rückkopplungen können dazu benutzt werden, aktiv Suchräume in anderen Bereichen eines Systems einzuschränken oder gar Vorhersagen für diese zu liefern. Dies ist nur dann sinnvoll, wenn verschiedene Module auf annähernd gleichen Ausschnitten einer Eingabe operieren.

- Neben diesen architektonischen Vorteilen (die sich vornehmlich auf die Performanz eines Systems auswirken)¹ ermöglicht die Einführung von Inkrementalität jedoch außerdem einen Sprung in der Qualität von Anwendungen: Erstmals sind dadurch Dialoganwendungen denkbar, die einen Benutzer unterbrechen können, etwa für den Fall, daß ein Widerspruch in der Eingabe entdeckt wird. Als zweites Beispiel sei erneut das Simultandolmetschen angeführt.

Alle hier angeführten Maßnahmen sind grundsätzlich nur dann denkbar, wenn ein modulares System mit kleinen Teilen der Eingabe zu operieren beginnt, es folglich inkrementell arbeitet.

Bisher ist der Aspekt der Inkrementalität hauptsächlich punktuell behandelt worden, so z.B. im Bereich der Generierung natürlicher Sprache [Fin96]. Zwei elaboriertere Ansätze finden sich bei Kitano [Kit94], der Übersetzungen auf der Basis von Beispielen mit Hilfe massiv-paralleler Rechner erzeugt, und bei INTARC, das zumindest bis zum Transfer eine inkrementelle Arbeitsweise zeigt [GSSW98].

Das hier vorgestellte System MILC (*Machine Interpreting with Layered Charts*) [Amt98] demonstriert, daß ein inkrementelles Vorgehen — in Abkehr zu bisher favorisierten Pipeline-Architekturen — möglich ist und in Performanzbereiche vordringt, die bis dato konventionellen Ansätzen vorbehalten waren.

2 Mehr-Ebenen-Charts

Die niedrigste Ebene der Repräsentation in einem System zur Verarbeitung gesprochener Sprache wird durch Worthypothesen gebildet, die angeben, mit welcher Wahrscheinlichkeit welches Wort in einem bestimmten Zeitintervall gesprochen wurde. In modernen Erkennungssystemen werden dafür Wortgraphen [AN95] verwendet, da sie dazu in der Lage sind, eine sehr große Menge an Äußerungshypothesen extrem kompakt darzustellen. So besteht der Wortgraph in Abb. 1 aus lediglich 461 Kanten², enthält jedoch $1,2 \cdot 10^{23}$ Pfade.

Diese offensichtliche Komplexität von Wortgraphen kann durch den Übergang zu Hypergraphen deutlich reduziert werden; in diesen ist eine Worthypothese mit mehreren Anfangs- und Endzeitpunkten versehen. Der Grund für die Effektivität dieses Übergangs liegt darin, daß Wortgraphen (zumal inkrementelle

¹Eine höhere Performanz geht im Bereich der Verarbeitung spontan gesprochener Sprache normalerweise mit einer besseren Qualität einher, da eine größere Anzahl von initialen Hypothesen über die gesprochenen Wörter berücksichtigt werden kann.

²Kanten repräsentieren möglicherweise gesprochene Wörter, während Knoten im Graphen mit Zeitpunkten assoziiert sind.

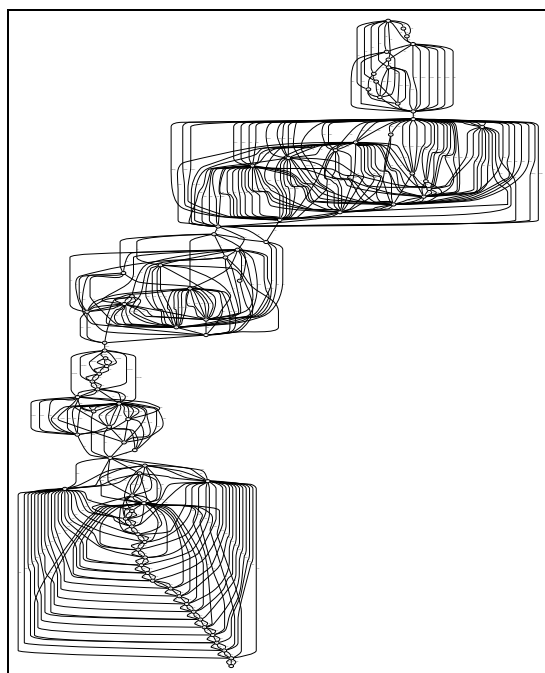


Abbildung 1: Ein Wortgraph

Wortgraphen) sehr viele ähnliche Worthypothesen enthalten, welche identische Wörter mit geringem zeitlichen Versatz präzisieren. Alleine durch den Wechsel zu Hypergraphen konnte die Geschwindigkeit einer partiellen syntaktischen Strukturanalyse um eine Größenordnung gesteigert werden.

Auf anderen Ebenen linguistischer Verarbeitung (vornehmlich im Bereich der syntaktischen Strukturanalyse und Generierung) sind Strukturen, die im wesentlichen isomorph zu Wortgraphen sind, zur Repräsentation partieller und vollständiger Resultate verwendet worden. Wir erweitern diese — Chart genannten Graphen [Kay80] — um die Hyperkanteneigenschaft und gelangen so zu Mehr-Ebenen-Charts, die dazu in der Lage sind, Hypothesen auf allen Ebenen linguistischer Verarbeitung integriert darzustellen.

Mehr-Ebenen-Charts sind gerichtete, azyklische Hypergraphen mit Knoten- und Kantenbezeichnungen. Knoten bezeichnen normalerweise Zeitpunkte, Hyperkanten repräsentieren vollständige oder unvollständige Resultate auf allen Ebenen innerhalb eines Systems. Abb. 2 zeigt den prinzipiellen Aufbau einer Mehr-Ebenen-Chart.

Diese Integriertheit der Darstellung hat weitreichende architektonische Vorteile:

- Der Austausch von Information zwischen Modulen wird wesentlich er-

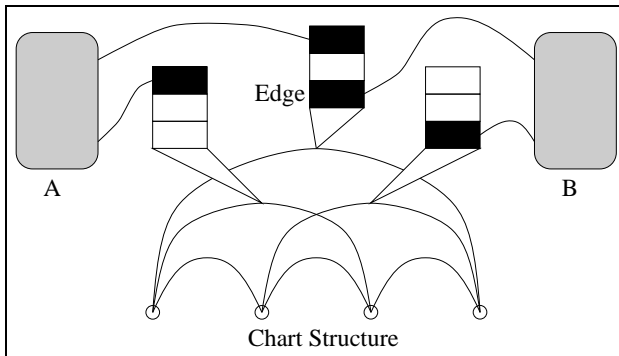


Abbildung 2: Der prinzipielle Aufbau einer Mehr-Ebenen-Chart

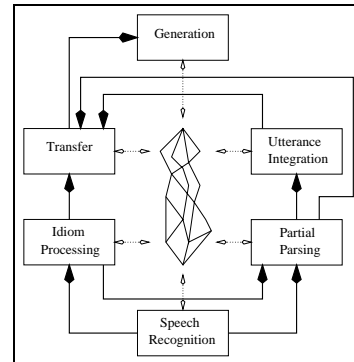


Abbildung 3: Die Architektur von MILC

leichtert, da alle Komponenten dieselbe Art der Darstellung benutzen. Anstatt für jede Schnittstelle zwischen zwei Komponenten von neuem den Prozeß und die Semantik des Datenaustausches zu definieren, kann jedwede Kommunikation auf den Transport von Hyperkanten reduziert werden.

- Eine Mehr-Ebenen-Chart kann verteilt gespeichert werden. Jede Komponente sieht jeweils nur den für sie relevanten Ausschnitt an Information, was einerseits die Komplexität von Komponenten reduziert und andererseits die Effizienz einer parallelen Implementation erhöht. Dezierte Informationskanäle sorgen für den Austausch von Daten zwischen Komponenten.
- Sämtliche partiellen Resultate auf allen Ebenen eines Systems sind sofort zugänglich. Dies ermöglicht es uns, das inkrementelle Paradigma bis in die Oberflächengenerierung beizubehalten und somit einzelne Teile einer Übersetzung auszugeben, sobald diese verfügbar sind.
- Obwohl in einem verteilten System prinzipiell kein globaler Zustand definiert ist, stellt die Vereinigung aller Kanten aller Komponenten jederzeit ein annähernd akkurates Bild des Fortschritts der Bearbeitung dar.

Neben der Integriertheit, die durch den Einsatz einer Mehr-Ebenen-Chart erreicht wird, bildet die Uniformität, die durch den Einsatz eines einzigen Formalismus zur Repräsentation linguistischen Wissens gesichert ist, eine weitere wichtige Eigenschaft von MILC. Wir adaptieren getypte, komplexe Merkmalstrukturen [Car92], eine moderne Sprache zur Beschreibung linguistischer Objekte. Die von uns durchgeführte Implementation verwendet zusammenhängende Speicherbereiche zur Erstellung von Merkmalstrukturen, so daß innerhalb

eines verteilten Systems Daten effizient ohne Verlust an Bedeutung transportiert werden können.

3 Architektur

Das System MILC demonstriert, wie die bisher geschilderten Prinzipien (Inkrementalität, Integriertheit und Uniformität) in einer Anwendung zur Übersetzung spontan gesprochener Sprache genutzt werden können, um einerseits ein performantes Gesamtsystem zu formen, und andererseits eine Experimentierplattform vorzuhalten, aufgrund derer weitergehende Architekturuntersuchungen vorgenommen werden können. Die globale Architektur des Systems ist in Abb. 3 zu sehen. Die graphenähnliche Struktur im Zentrum der Abbildung symbolisiert den Einsatz einer Mehr-Ebenen-Chart für alle Komponenten. Kästen stellen individuelle Module dar, gerichtete Pfeile repräsentieren Kommunikationskanäle.

Die erste Komponente ist ein HMM-basierter Worterkenner [HJH96], der inkrementell einen Wortgraphen konstruiert. Dieser wird online in einen Hypergraphen konvertiert, dessen Kanten an die Idiomererkennung und die syntaktische Verarbeitung verteilt werden.

Die Idiomererkennung sucht nach festen Redewendungen, die starr lexikalisiert sind. Beispiele für derartige Konstruktionen, deren Semantik i.A. nicht-kompositional ist, sind etwa Begrüßungen (*guten Tag*) oder Idiome, die den Dialogfluß fortsetzen (*Einen Moment, bitte*). Aufgrund des feststehenden Charakters der Redewendungen kann eine lexikalisch orientierte, inkrementelle Suche auf der Grundlage von Oberflächenformen stattfinden, ohne auf komplexere syntaktische Mechanismen zurückgreifen zu müssen. Ist ein Idiom gefunden, wird dies, zusammen mit einer Merkmalstruktur, die es beschreibt, an die syntaktische Strukturanalyse weitergegeben. Gleichzeitig werden die Worthypothesen, die zur Konstruktion benutzt wurden, mit einer Strafe versehen, die deren weitere Inkorporierung in Konstituenten anderer Art unwahrscheinlicher macht. Der Grund dafür liegt in der Annahme, daß die Idiomererkennung eine höhere Sicherheit aufweist, da sie auf exakten lexikalischen Beispielen arbeitet. Allerdings werden die zugrundeliegenden Worthypothesen nicht vollständig entfernt, so daß bei Bedarf weiterhin auf sie zugegriffen werden kann.

Die Komponente zur Idiomererkennung ist gleichzeitig ein Beispiel für die Art und Weise, in der zusätzliche Wissensquellen und Module in MILC integriert werden können. Durch die Benutzung einer Mehr-Ebenen-Chart ist eine einheitliche Bewertung von Hypothesen auf allen Ebenen möglich; durch die Modifikation von Bewertungen und deren Vererbung über Modulgrenzen hinweg kann eine Komponente Einfluß auf Operationen innerhalb anderer Module

nehmen.

Die syntaktisch/semantische Verarbeitung von Äußerungen ist zweigeteilt: Eine Komponente erstellt partielle Strukturanalysen, die andere integriert diese in größere Zusammenhänge. Der Grund hierfür liegt zum einen darin, daß Spontansprache gehäuft Phänomene aufweist, die als ungrammatisch im Sinne einer Standard-Grammatik angesehen werden müssen, die aber dennoch verarbeitet werden sollen. Demzufolge liegt das Ziel für die erste Komponente darin, "kleine" Konstituenten zu finden — z.B. Nominalphrasen und Präpositionalphrasen —, ohne auf Kohärenz über einen weiteren Ausschnitt zu achten. Die zweite Ursache ist, daß Verbalphrasen mit Verbendstellung, wie im Deutschen Nebensatz, die Konstruktion von Komplementkomplexen unnötig teuer machen, wenn diese strikt von links nach rechts stattfindet. Stattdessen ist es wesentlich ökonomischer, erst auf das Eintreffen des Verbs zu reagieren, und dann die Verbalphrasen und Satzthesen zu erstellen. Dies erfordert allerdings eine leichte Aufweichung vom stark inkrementellen Paradigma, indem potentiell Komplemente links vom Verb gesucht werden müssen.

Die Transferphase übersetzt deutsche semantische Repräsentationen in ihre englischen Entsprechungen. Dieses Modul ist — wie alle anderen innerhalb von MILC — inkrementell ausgelegt und benutzt einen Chart-basierten Ansatz [Amt97]. Der Transfer beginnt mit kleinen Fragmenten semantischer Beschreibungen, die von der Strukturanalyse oder der Idiomverarbeitung geliefert werden, und integriert diese nach und nach in größere Konstrukte. Bei deren Aufbau kann dank der Chart-Eigenschaft auf bereits transferierte Elemente zurückgegriffen werden.

Die Generierung stellt die letzte Komponente dar, die englische Oberflächenrepräsentationen aus englischen semantischen Beschreibungen erstellt. Die zeitliche Ordnung, in der die partiellen Generierungsergebnisse erstellt werden, entspricht im wesentlichen der Ordnung der Eingabe. MILC ist dadurch in der Lage, sukzessiv längere Fragmente zu generieren, wodurch der inkrementelle Charakter des Systems deutlich gemacht wird.

Wir fordern nicht, daß eine einzelne Analyse die gesamte Eingabe überspannt; dies wäre angesichts der Eigenschaften von Spontansprache (Häitationen, Abbrüche, Neuanfänge, etc.) auch aussichtslos. Stattdessen werden auf allen Ebenen erfolgversprechende Analysefragmente gesammelt und weitergegeben. Innerhalb der Generierung wird ständig der beste Pfad durch den Graphen englischer Äußerungsfragmente vorgehalten. Ändert sich dieser Pfad — entweder durch das Hinzukommen weiterer Satzglieder oder durch Neubewertung bereits bestehender Anteile — so wird der komplette Pfad neu ausgegeben.

4 Experimente

Das System MILC wurde mit fünf Dialogen aus der Domäne der Terminabsprache (dem Verbmobil-Korpus) getestet; zwei dieser Dialoge wurden zur Entwicklung der Grammatiken verwendet, während die restlichen Dialoge als Testset dienten, und lediglich die notwendigen Lexika ergänzt wurden. Die Länge einer Äußerung betrug im Schnitt 20 Wörter oder 6,5 Sekunden. Initial bestand ein Wortgraph aus 6050 Worthypothesen und enthielt $1,5 \cdot 10^{213}$ Pfade. Die Worterkennungsrate betrug etwa 70%. Die Laufzeit für die komplette Übersetzung einer Äußerung betrug ca. 40 Sekunden auf einer Doppelprozessor-UltraSparc 4 (1GB), was ungefähr sechsfacher Echtzeit entspricht. Die Übersetzungen wurden von englischen Muttersprachlern evaluiert. Dies ergab eine Erfolgsrate von 60% approximativ korrekten Übersetzungen³.

Zwei Aspekte der Experimente scheinen wesentlich. Durch den durchgängig inkrementellen Charakter des Systems ist es erstmals möglich, bereits Fragmente einer übersetzten Äußerung auszugeben. Dadurch kann der zugrundeliegende Prozeß direkt visualisiert werden; außerdem muß im Prinzip nicht auf das Ende der Eingabeäußerung gewartet werden, bevor Teile einer englischen Übersetzung produziert werden. Gleichzeitig ergibt sich damit allerdings eine recht abgehackte zielsprachliche Realisierung, wie das folgende Fragment in Tab. 1 zeigt:

Tabelle 1: Einige Ausgabehypothesen von MILC

me Yes it I would suggest an appointment we the fifth
me Yes it I would suggest an appointment on friday we the fifth
me Yes it I would suggest an appointment on friday we something the fourteenth
me Yes it I would suggest an appointment on friday we from the fourteenth

Es ist nicht sonderlich überraschend, daß eine weitergehende Bearbeitung der Fragmente notwendig ist. Ein möglicher Ansatz ist hier, die englischen Äußerungen teilweise zu reanalysieren, um deren Qualität zu erhöhen.

Weiterhin ergab ein Vergleich zwischen der inkrementellen Implementation und einer äquivalenten, nicht-inkrementellen Fassung, daß die inkrementelle Version lediglich um 17% weniger effizient ist. Dies zeigt, daß inkrementelle Verfahren durchaus in Bereiche vordringen können, die bislang konventionellen Methoden vorbehalten waren.

³Approximativ korrekt meint, daß das wesentliche Ziel der Äußerung (Vorschlag etc.) sowie zentrale Inhalte korrekt übersetzt wurden.

5 Zusammenfassung

Die Konstruktion inkrementeller Systeme im Rahmen der automatischen Verarbeitung natürlicher Sprache erscheint der logische nächste Schritt auf dem Wege zu hochwertigen, natürlich wirkenden Anwendungen. Das hier vorgestellte System MILC demonstriert, daß eine inkrementelle Übersetzung spontan gesprochener Sprache im Bereich des Möglichen liegt, und stellt zugleich eine Experimentierplattform für architektonische Untersuchungen bereit. Die Anwendung der Architekturprinzipien Integriertheit und Uniformität sorgt für einen sauberen, erweiterbaren Entwurf und erleichtert etliche softwaretechnische Aspekte innerhalb eines Systems.

Das System wurde anhand einiger spontansprachlicher Dialoge getestet und erreicht 60% approximativ korrekte Übersetzungen in etwa sechsfacher Realzeit.

Dies demonstriert, daß entgegen bisheriger Erfahrungen die Einführung inkrementeller Algorithmen nicht von vorneherein aus Effizienzgründen scheitern muß. Wir erwarten vielmehr, daß zukünftige Entwicklungen weit stärker als bisher Nutzen aus den architektonischen Vorteilen inkrementeller Systeme ziehen werden.

Literatur

- [Amt97] Jan W. Amtrup. Perspectives for Incremental MT with Charts. In Christa Hauenschild and Susanne Heizmann, editors, *Machine Translation and Translation Theory. Perspectives of Co-operation*, Text, Translation, Computational Processing (TTCP), number 1. Mouton de Gruyter, 1997.
- [Amt98] Jan W. Amtrup. *Maschinelles Dolmetschen mit Mehr-Ebenen-Charts*. PhD thesis, Universität Hamburg, 1998.
- [AN95] Xavier Aubert and Hermann Ney. Large Vocabulary Continuous Speech Recognition Using Word Graphs. In *ICASSP 95*, 1995.
- [Car92] Bob Carpenter. *The Logic of Typed Feature Structures*. Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, 1992.
- [Fin96] Wolfgang Finkler. *Automatische Selbstkorrektur bei der inkrementellen Generierung gesprochener Sprache unter Realzeitbedingungen — Ein empirisch-simulativer Ansatz unter Verwendung eines*

- Begründungsverwaltungssysteme*. PhD thesis, Universität des Saarlandes, Saarbrücken, 1996.
- [GSSW98] Günther Görz, Jörg Spilker, Volker Strom, and Hans Weber. Architectural Considerations for Conversational Systems – The Verbomobil/INTARC Experience. In Y. Wilks, editor, *Machine Conversions*. Kluwer, Dordrecht, 1998.
- [HJH96] Kai Huebener, Uwe Jost, and Henrik Heine. Speech Recognition for Spontaneously Spoken German Dialogs. In *ICSLP96*, Philadelphia, 1996.
- [Kay80] Martin Kay. Algorithmic Schemata and Data Structures in Syntactic Processing. Technical Report CSL-80-12, Xerox Palo Alto Research Center, Palo Alto, 1980.
- [Kit94] Hiroaki Kitano. *Speech-to-Speech Translation: A Massively Parallel Memory-Based Approach*. Kluwer Academic Publishers, Boston, 1994.
- [MWT80] William D. Marslen-Wilson and Lorraine K. Tyler. The Temporal Structure of Spoken Language Understanding. *Cognition*, 8:1–71, 1980.
- [Niv93] Michael Niv. *A Computational Model of Syntactic Processing: Ambiguity Resolution from Interpretation*. PhD thesis, Univ. of Pennsylvania, 1993.



Jan W. Amtrup studierte Informatik an der Universität Hamburg, Diplom 1992. Von Mitte 1993 bis Anfang 1998 war er wissenschaftlicher Mitarbeiter an der Universität Hamburg. Im Rahmen des BMBF-Verbundprojektes Verbomobil beschäftigte er sich mit der Architektur des Transfers und mit infrastrukturellen Aspekten. Seit April 1998 ist er Research Analyst/Computer Specialist am Computing Research Lab (CRL) der New Mexico State University. Dort arbeitet er vorwiegend an der Architektur und dem Design von Übersetzungssystemen.