

# **Die Implementierung japanischer Sprache in Head-Driven Phrase Structure Grammar (HPSG)**

–

## **und was wir damit gemacht haben**

Melanie Siegel



# Motivation

- **Anwendungen** für computerlinguistische Verfahren und Analysen werden zunehmend realisierbar und eingesetzt.

# Maschinelle Übersetzung

**SDL** | FreeTranslation.com

Professionelle Übersetzung Übersetzer Tools Sprache lernen

Ausgangssprache Englisch



Zielsprache Japanisch

Übersetzen

Robin has read a book.  
The cat is in the house.



Wir unterstützen .doc(x), .ppt(x), .odt, .pdf & .txt (Maximale Dateigröße 5 MB)

📁 Datei(en) auswählen

ロビンの本を読みました。  
猫は家こいます。



# Automatische Fragebeantwortung

**START**  
Natural Language Question Answering System

How many inhabitants does Darmstadt have? [Ask Question >](#)

=> How many inhabitants does Darmstadt have?

1313.0 people live in Darmstadt town, Vanderburgh County, Indiana.

Source: U.S. Census Bureau

[Back to home page](#)

<http://start.csail.mit.edu/index.php>

# Dialogsysteme



## A. L. I. C. E. Artificial Intelligence Foundation

Promoting the development and adoption of **ALICE** and **AIML Free Software**

A.L.I.C.E. Silver Edition

DAVE E.S.L. bot

C.L.A.U.D.I.O Personality Test

GET SITEPAL AVATARS

Get your own free chat bot: *Be Your Own Botmaster*, 2nd ed.



Click here to chat with Talking Animate



Get **ALICE Showcase Edition** with talking **Sitepal** avatar.



**CallMom ALICE 2.0** Android app



### A.L.I.C.E. and judge

**You said:** Oh, how nice!

**A.L.I.C.E.:** Yes it is.

You say:

Say

*Get an A.I. Foundation membership certificate.*

### Conversation Log:

judge: Hi Alice  
ALICE: Hi there!  
judge: what will be the weather today?  
ALICE: I cannot predict the future.  
judge: how's the weather at your place?  
ALICE: Cloudy.  
judge: Where are you?  
ALICE: California.  
judge: Oh, how nice!  
ALICE: Yes it is.

<http://alice.pandorabots.com/>

Melanie Siegel

# Grammatik-, Terminologie- und Stilprüfung

The screenshot shows the Epic Editor interface with a document titled "ErsteSchritteEP.xml". The document content is as follows:

**Benutzerhandbuch**

**1. Erste Schritte**

**1.1. Installieren der SIM-Karte, des Akkus und der Speicherkarte**

Bewahren Sie alle SIM-Karten außerhalb der Reichweite von **kleinen Kinder** auf.

Wenden Sie sich bezüglich Verfügbarkeit und weiterer Informationen über die Verwendung der **Service Provider**, der Netzbetreiber oder ein Drittanbieter sein.

Verwenden Sie nur kompatible **MMC-Speicherkarten** (MultiMediaCard) mit diesem Gerät. **Kartensteckplatz** und sind mit diesem Gerät nicht **Gerät beschädigt** werden. Außerdem können die a

1. Bevor Sie den Akku **herausnehmen**, mü
2. Legen Sie die SIM-Karte in den **SIM-I**

A context menu is open over the text "MMC-Speicherkarten" with the following options:

- Gesperrter Tem...
- Ersetzen durch :
  - MMC-Karten**
- Fehler bearbeiten
- Fehler ignorieren
- Step-through
- Vorherige Markierung
- Nächste Markierung

# Motivation

- größere Komplexität der Anwendungen
- Anforderungen an **Präzision** und den **Informationsgehalt** der linguistischen Analyse steigen
- aber nicht auf Kosten von **Robustheit** und **Effizienz**

# Motivation

- **Anforderungen:**
  - Reichhaltige Information
  - präzise Information
  - Wohldefinierte Ausgabestrukturen
  - Robustheit
  - Große Abdeckung
  - Große und erweiterbare Lexika
  - Interfaces zur Vorverarbeitung
  - Übertragbarkeit der Ansätze auf weitere Sprachen
  - Effiziente Verarbeitung
- *Die JACY-Grammatik wurde für Anwendungen konzipiert und implementiert und in komplexen Anwendungen eingesetzt.*

# Die JACY Grammatik: Geschichte

- 1998-2000
  - Verbmobil: Maschinelle Übersetzung anwendungsorientierter gesprochener Dialoge  
<http://verbmobil.dfki.de/>
- 2001-2002
  - Kooperation mit YY Technologies (CA, USA): Automatische Email-Beantwortung (Kooperation mit Stephan Oepen, Ulrich Callmeier, Mönique Sugimoto, Atsuko Shimada, Dan Flickinger)  
<http://www.dfki.de/~siegel/jacy/jacy.html>
- 2002-2004
  - EU-Projekt DEEP THOUGHT: Hybride und flache Methoden für wissensintensive Informationsextraktion (Kooperation mit Stephan Oepen, Ulrich Callmeier, Ann Copestake, John Carroll, Dan Flickinger und anderen)  
<http://www.project-deepphought.net>
- 2005
  - Industriauftrag der japanischen Firma NTT, Ontologieextraktion (Kooperation mit Francis Bond)

# Framework

- Head-Driven Phrase Structure Grammar (HPSG)
  - Attribut-Wert-Strukturen
  - Typenhierarchie
  - Effiziente Verarbeitung
- Minimal Recursion Semantics (MRS)
  - flacher semantischer Formalismus
  - Zusammenspiel mit getypten Attribut-Wert-Strukturen
  - unterspezifizierte Strukturen
  - auflösbar zu Robust Minimal Recursion Semantics (RMRS)

# Die Werkzeuge

- **LKB**: System für Grammatikentwicklung ([Copestake 2002](#))
- **PET**: effizientes Verarbeitungssystem für HPSG-Grammatiken ([Callmeier 2000](#))
- **ChaSen**: Tokenisierer und POS-Tagger ([Asahara & Matsumoto 2000](#))
- **[incr tsdb]**: Werkzeug zum Testen von Grammatiken ([Oepen & Carroll 2000](#))
- **Heart-of-Gold**: Architektur für die Kombination von tiefer und flacher Verarbeitung ([Callmeier, Eisele, Schäfer & Siegel 2004](#))

# Open Source

- JACY ist eine Open-Source Grammatik für Japanisch.
- JACY Homepage:  
<http://www.delph-in.net/jacy>

# Multilinguale Grammatikentwicklung

- HPSG-Grammatiken im Delph-In Kontext:
  - Deutsch (50.000 Lexikoneinträge)
  - Englisch (12.300 Lexikoneinträge)
  - Japanisch (35.000 Lexikoneinträge)
  - Norwegisch (84.240 Lexikoneinträge)
  - Italienisch (4.850 Lexikoneinträge)
  - Portugiesisch
  - Griechisch
- Grammar Matrix: Neue Grammatiken können schnell implementiert werden
- RMRS als gemeinsamer Semantikformalismus: Nutzung für Anwendungen

# Computerlinguistische Aufgaben / Japanisch

ロビンが本を読んだ。

- Tokenisierung:

ロビン が 本 を 読んだ 。

↑        ↑        ↖        ↙

Katakana Kanji Hiragana Punkt\_Satzende

- Part-of-Speech Tagging:

ロビン が 本 を 読んだ 。

↑        ↑        ↙

Name Kasuspartikel Verb

# Computerlinguistische Aufgaben / Japanisch

- Morphologie-Analyse:

ロビン が 本 を 読んだ 。

↑  
lemma: yomu  
tense: past  
honorific: plain

- Namenserkennung:

ロビン が 本 を 読んだ 。

↑  
name: Robin

# Computerlinguistische Aufgaben / Japanisch

- Chunk Parsing:

ロビン が 本 を 読んだ 。

——— ——— ———  
PP      PP      VP

- Statistisches Parsing:

read(Robin,book)

# Computerlinguistische Aufgaben / Japanisch

- Grammatische Analyse/Semantische Analyse

```

TEXT  ロビンが本を読んだ
TOP   h1

RELS  {
        | proposition_m_rel | named_rel | def_rel | _hon_n | udef_rel | _yomu_v
        | LBL h1 | LBL h5 | LBL h8 | LBL h12 | LBL h15 | LBL h19
        | ARG0 e2 tense=past | ARG0 x6 | ARG0 x6 | LBL h12 | ARG0 x13 | ARG0 e2 tense=past
        | MARG h3 | CARG ROBIN | RSTR h10 | ARG0 x13 | RSTR h16 | ARG1 x6
        | | | BODY h11 | | BODY h18 | ARG2 x13
    }

HCONS {h3 qeq h19, h10 qeq h5, h16 qeq h12}
ING   {}

```

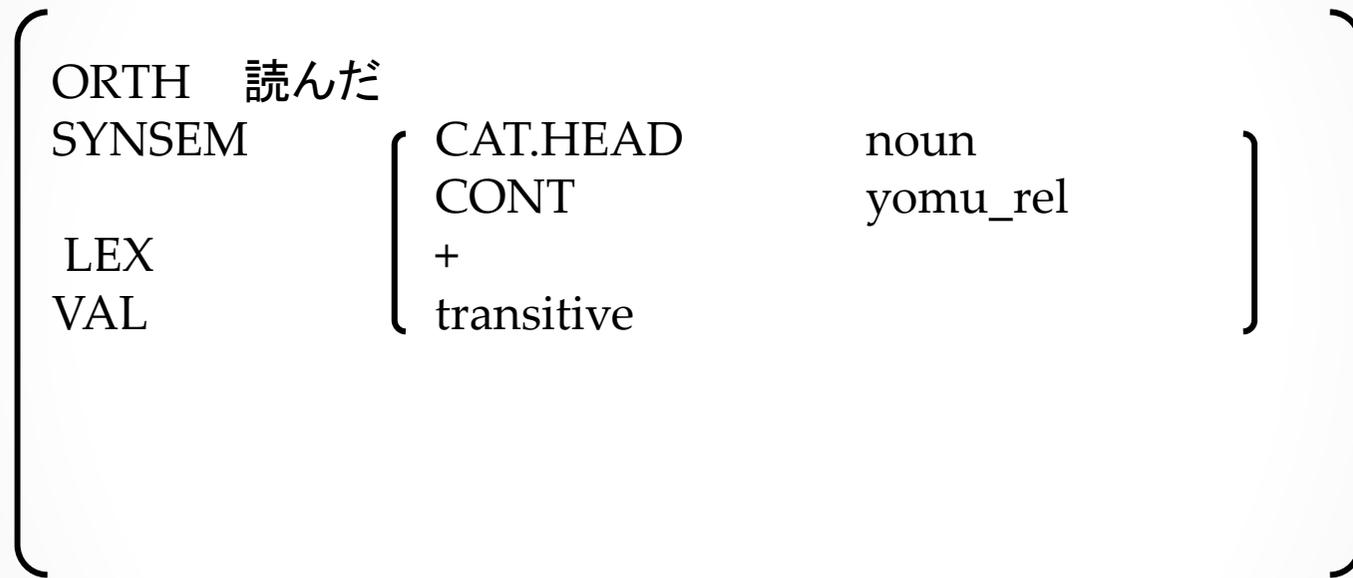
[Show XML source](#)

- Nullpronomina, Honorifikation, Topikalisierung, Wortstellung, Subkategorisierung...

# Grundbegriff: Zeichen

- Zeichen:
  - Komplexe Attribut-Wert-Struktur
  - Information von verschiedenen linguistischen Ebenen einer Phrase oder einer lexikalischen Einheit
  - Information:
    - Orthographische Information in ORTH
    - Syntaktische und semantische Information in SYNSEM
    - Lexikalischer Status in LEX
    - Nichtlokale Information in NON-LOCAL
    - Kopfinformation, die im Baum hochgereicht wird in HEAD
    - Information über Subkategorisierung in VAL

# Grundbegriff: Zeichen



# Grundbegriff : Typen

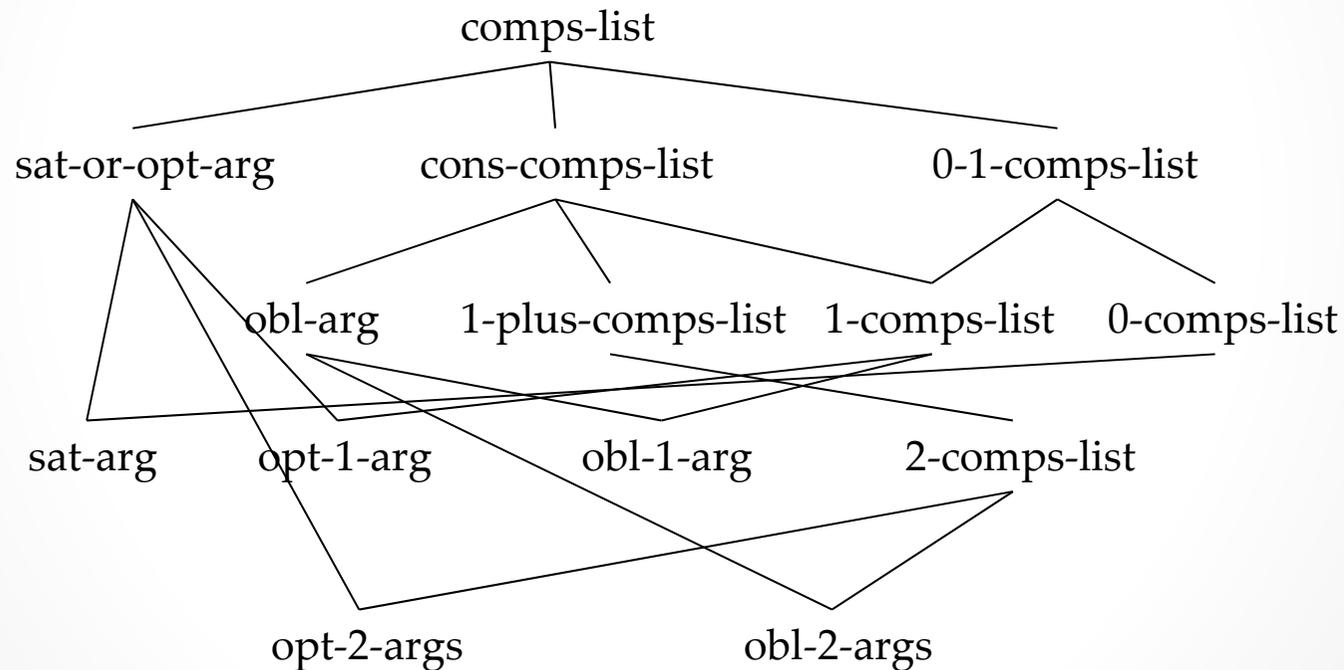
- Die Grammatikimplementierung basiert auf einem Typsystem.
- 990 lexikalische Typen für die syntaktischen, semantischen und pragmatischen Eigenschaften
- 188 Typen für die Eigenschaften der Phrasen, lexikalischen Regeln und der Flexions- und Derivationsregeln
- 50 Regeln für Flexions- und Derivationsmorphologie und lexikalische Regeln
- 47 Phrasenstrukturregeln

# Subkategorisierung in der japanischen Grammatik

- Nullpronomina
  - „*yonda*“ – „gelesen“ ist ein perfekter japanischer Satz, wenn Subjekt und Objekt sich aus dem Kontext ergeben.
- Wortstellungsvariation
  - *moritsu ga hon wo yonda* (Moritz hat ein Buch gelesen)
  - *hon wo moritsu ga yonda* (Ein Buch hat Moritz gelesen)
- Topikalisierung
  - *robin wa yonda.* (Robin hat gelesen)
  - *hon wa yonda.* (Ein Buch hat (er) gelesen)
- „Doppel-Subjekte“
  - *zoo ga hana ga nagai* (Elefanten haben große Nasen)

# Subkategorisierung in der japanischen Grammatik

- Typhierarchie der Komplementlisten



# Lexikalische Darstellung

- Valenz-Struktur eines typischen transitiven Verbs:

VAL  $\left( \begin{array}{l} \text{ga-wo-transitive} \\ \text{UNSAT plus} \\ \text{SUBJ opt-1-arg \&lt; [LOCAL.CAT.HEAD.CASE ga]\>} \\ \text{COMPS opt-1-arg \&lt; [LOCAL.CAT.HEAD.CASE wo]\>} \\ \text{SPR 0-comps-list} \end{array} \right)$

# Vom Stamm zum Wort

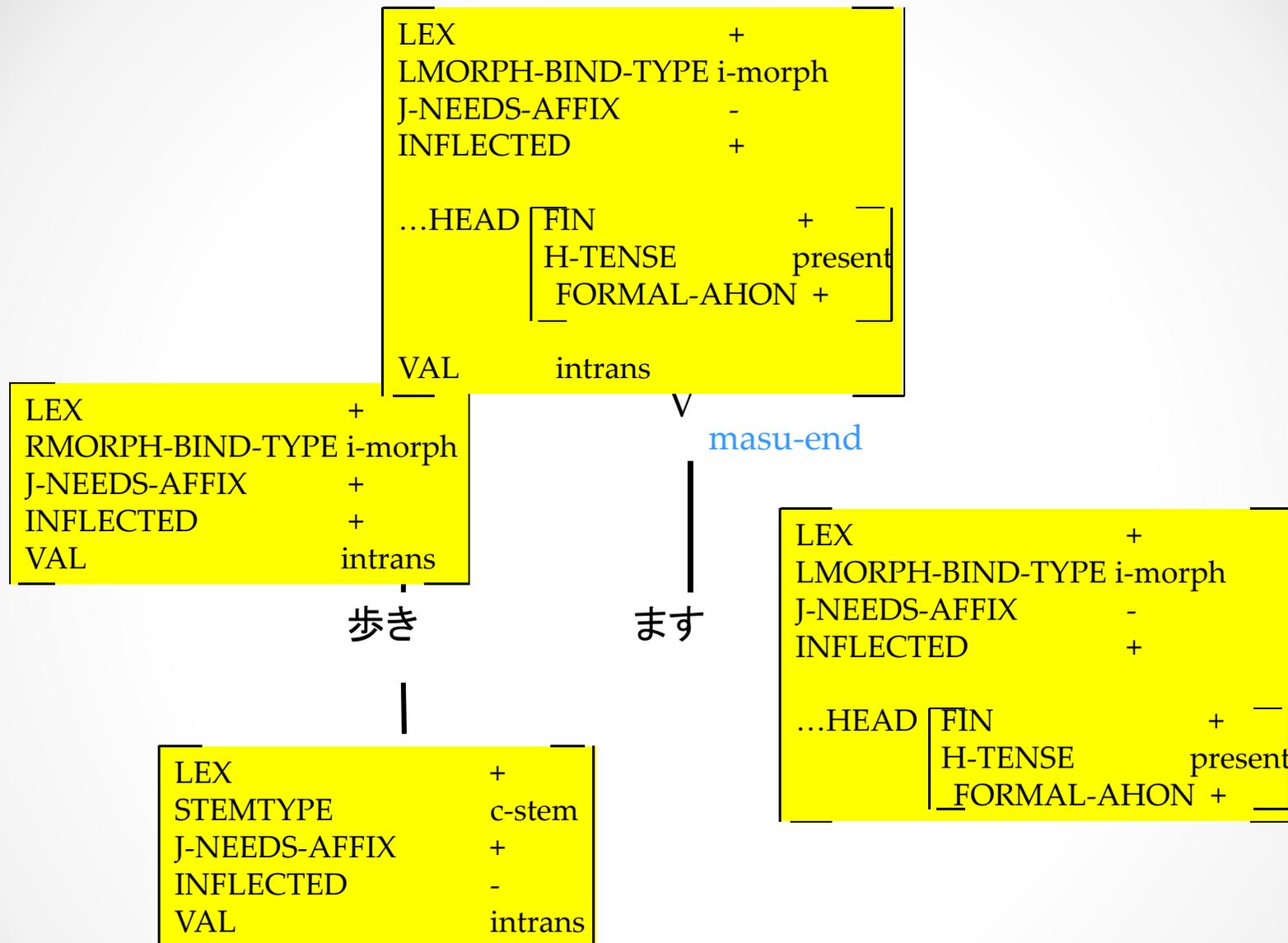
歩き ます  
aruki masu

- Wortstämme
  - kategorisiert nach morphologischem und Valenzverhalten
  - in einer Typhierarchie organisiert
- Flexionsregeln werden auf Stämme angewendet.
  - verursachen morphologische Veränderungen und geben dem Ergebnis einen morphologischen Typ
- Regeln, die Stämme zu Wörtern machen
- Derivationsregeln

# Endungen

歩き ます  
aruki masu

- Verbendungen werden in der Grammatik mit einer binären Regel mit dem Verb verbunden
  - (*vstem-vend*, eine Instanz von *head-specifier*)
- Sie fügen Informationen über adressaten-bezogene Honorifikation, Tempus, Modus usw. hinzu
- Argumentstruktur des Stamm-Endungskomplexes kommt vom Stamm.
- Endung subkategorisiert für den Stamm.



# Auxiliare

- treten im Kombination mit anderen Verben auf
    - Aspektinformation
    - Perspektivinformation
    - Information über Honorifikation
  - Information über Subkategorisierung ist eine Kombination der SUBCAT-Information von Verb und Auxiliar
- *head-specifier-rule*.

# Auxiliartypen

- Aspekt-Auxiliare.
  - fügen Aspekt-Information hinzu
  - Z.B., いる、ある (*iru, aru*)
- Perspektiv-Auxiliare.
  - Fügen ein mit *ni* (Dative) markiertes Argument in die Argumentstruktur des Prädikats ein.
  - Z.B., くれる (*kureru*)
  - Als Subjekt-Kontrollverben behandelt.
- Obj-id-Auxiliare.
  - Z.B., もらう (*morau*)
  - Etablieren eine Kontrollrelation zwischen mit Dativ-Objekt und dem eingebetteten Subjekt.

# Aspekt-Auxiliare

- Reiner Aspekt:

ケ-キ	を	食べて	いる
<i>keeki</i>	<i>wo</i>	<i>tabete</i>	<i>iru</i>
<i>Kuchen</i>	<i>ACC</i>	<i>essen</i>	<i>AUX</i> <i>(progressiv)</i>

Ich esse gerade Kuchen

- Fügen nur die Aspektinformation in die Semantik des Satzes ein.

- Aspekt:

ケ-キ	が	食べて	ある
<i>keeki</i>	<i>ga</i>	<i>tabete</i>	<i>aru</i>
<i>Kuchen</i>	<i>NOM</i>	<i>essen</i>	<i>AUX</i> <i>(perfektiv)</i>

Der Kuchen ist gegessen.

- Ändern die Valenz des Verbkomplexes.

# Aspekt-Auxiliare

- Auxiliare mit komplexem Aspekt:

ロビン	が	ケ-キ	を	食べて	見る
<i>Robin</i>	<i>ga</i>	<i>keeki</i>	<i>wo</i>	<i>tabete</i>	<i>miru</i>
<i>Robin</i>	<i>NOM</i>	<i>Kuchen</i>	<i>ACC</i>	<i>essen</i>	<i>AUX (modal: versuchen)</i>

Robin versucht, den Kuchen zu essen.

- Fügen der Semantik eine Relation hinzu.
- *Ihr ARG1 ist mit dem ARG1 des Verbs identisch und ihr ARG2 ist der Handle einer Proposition, die Skopus über die Relation des Verbs hat.*

# Perspektiv-Auxiliare

先生 が 私 に 本 を 買って くれた

*sensei ga watashi ni hon wo katte kureta*

*Lehrer NOM mit DAT Buch ACC kaufen AUX (subj-control)*

Die Lehrerin hat mir ein Buch gekauft.

- ARG1 des Auxiliars ist identisch mit dem ARG1 des Verbs.
- Ein Dativobjekt als ARG2 wird hinzugefügt.
- Die Proposition über dem Verb wird mit dem ARG3 des Auxiliars verbunden.

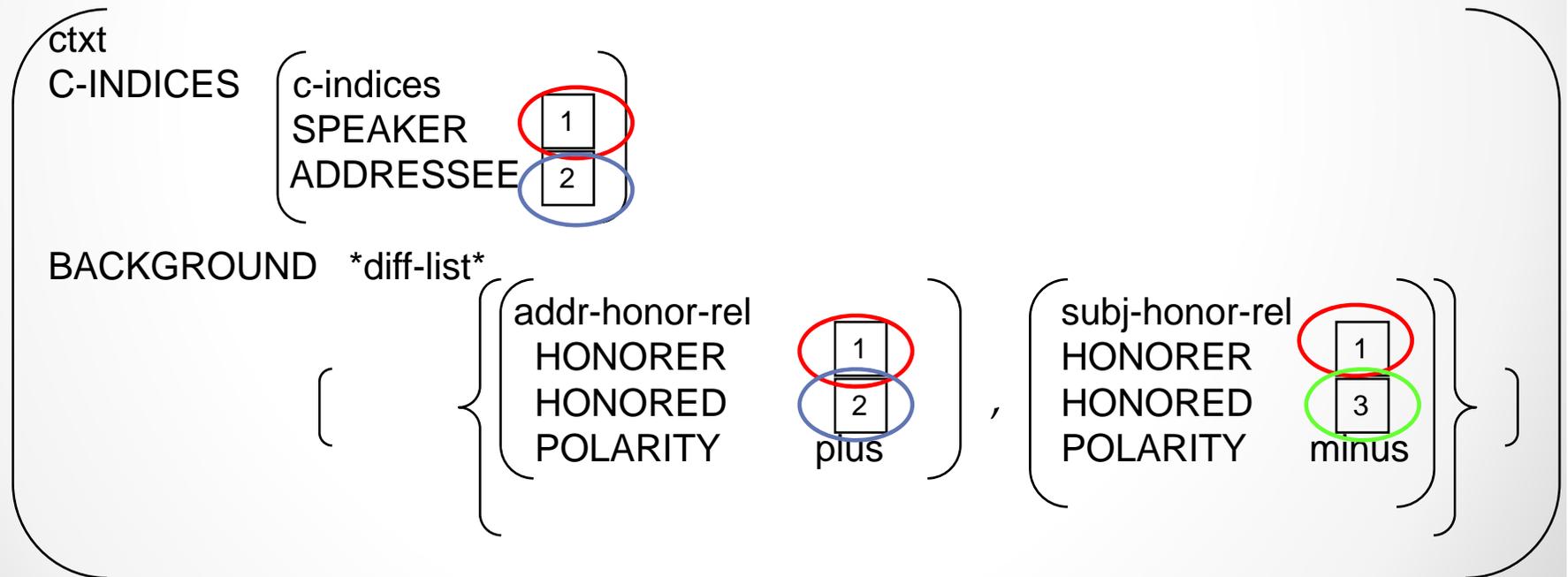
# Perspektiv-Auxiliare



# Pragmatische Information steht in CONTEXT

z.B. 食べております

(ich) esse



# JACY MRS-Semantik

- Grammatikentwicklung in einem multilingualen Kontext
- → Schwerpunkt auf **parallele und konsistente semantische Repräsentationen**
- Der größte Teil der semantischen Repräsentation in der englischen Grammatik konnte direkt auf das Japanische übertragen werden.
- Besondere Behandlung brauchten z.B.:
  - Nominalisierungen und Verbalnomen
  - Numerale Klassifikatoren
  - Relativsätze und Adjektive

# Hybride Sprachverarbeitung

- Die Analyseergebnisse von NLP-Tools auf flacheren Verarbeitungsebenen werden von Komponenten auf tieferen Ebenen genutzt.
  - Z.B.: Die tiefe linguistische Analysekomponente (mit der HPSG-Grammatik) nutzt Default-Lexikoneinträge für Namen, die die flache Analysekomponente (Namenserkenkung) Sprout liefert.

New text System status System log

成田に住んでいます Depth: 110 Analyze

xmltext ChaSen SProUTpiXML Sprout SproutFS PET RmrsMerge result

```

TEXT 成田に住んでいます。
TOP h1
RELS {
  | named_rel | def_rel | _ni_p | | | | | |
  | LBL h3 | LBL h6 | LBL h1 |
  | ARG0 x4 | ARG0 e10 tense=u | _sumu_v |
  | CARG NARITA_1 | RST h8 | ARG1 x4 | ARG0 e2 tense=present | proposition_m_rel |
  | | BODY h9 | ARG2 e2 tense=present | MARG h15 | _iru_v |
  | | | | | LBL h10001 | ARG0 e2 tense=present | LBL h10002 | LBL h17 |
  | | | | | | | ARG0 e2 tense=present | ARG0 e2 tense=present |
}
HCONS {h8 qeq h3, h15 qeq h17}
ING {h1 ing h10001, h1 ing h10002}
    
```

Show XML source

# Hybride Sprachverarbeitung

- Die Analyseer-  
Verarbeitungs-  
Ebenen genau
- Zum Beispiel  
(PET) Default  
der POS Tag



auf flacheren  
Komponenten auf tieferen

Analyse der HPSG-Grammatik  
HPSG-Speech Kategorien, die



xmltext ChaSen SproutFS Sprout SProUTpiXML PET result

```

TEXT 茶せんは何ですか
TOP  h1
RELS  {
    generic_noun | udef_rel | _wa_p | _thing_n | wh_rel | cop_id_rel | question_m_rel
    LBL          | LBL h7   | LBL          | LBL h16  | LBL h20   | LBL          | LBL          |
    LBL h3       | ARG0 x4   | ARG0 e13 tense=u | LBL h16  | ARG0 x19  | LBL          | LBL h1       |
    ARG0 x4      | RSTR h11  | ARG1         | ARG0 x19  | RSTR h24  | ARG0 e2 tense=present | ARG0 h31    |
    BODY h10     | ARG2 e2 tense=present |
}
HCONS {h11 qeq h3, h24 qeq h16, h31 qeq h25}
ING   {}
    
```

Melanie Siegel

[Show XML source](#)

# Hybride Sprachverarbeitung

Liefere das tiefste mögliche Ergebnis.

(Dabei sind aber alle anderen Ergebnisse verfügbar.)

**Heart of Gold - Online Demo**

[New text](#) [System status](#) [System log](#)  
モモくんが電話した Depth: 110 Analyze

xmltext ChaSen SProUTpiXML Sprout SproutFS PET RmrsMerge **result**

```
TEXT  モモくんが電話した。
TOP   h1

      |proposition_m_rel| |_momo_n| |_kun_n| |_denwa_s|
      |LBL h1| |LBL h5| |LBL h12| |LBL h19|
      |ARG0 e2 tense=past| |ARG0 x6| |ARG0 x13| |ARG0 e2 tense=past|
      |MARG h3| |RSTR h9| |RSTR h16| |ARG1 x13|
RELS  { |BODY h11| |BODY h17| |ARG2 x6| }
      |_denwa_s|
      |LBL h19|
      |ARG0 e2 tense=past|
      |ARG1 x13|

HCONS {h3 qeq h19, h9 qeq h5, h16 qeq h12}
ING   {h5 ing h10001, h5 ing h10002}

Show XML source
```

Melanie Siegel

# Hybride Sprachverarbeitung

Kombination von Modulen und Grammatiken für verschiedene Sprache. Jede Sprache hat ihre eigene Konfiguration von Modulen und Grammatiken.

The screenshot shows a web-based interface titled "Heart of Gold Settings". It is divided into two main sections: "Heart of Gold Settings:" and "Input text:".

**Heart of Gold Settings:**

- Depth:** A vertical list of radio buttons with values 10, 20, 30, 50, 80, and 100. The 100 option is selected.
- Languages:** A vertical list of radio buttons with language names and corresponding national flags:
  - English (UK flag) - selected
  - German (German flag)
  - Italian (Italian flag)
  - Norwegian (Norwegian flag)
  - Japanese (Japanese flag)
  - Greek (Greek flag)

**Input text:** A large, empty rectangular text input field.

At the bottom right of the settings area, there are four buttons: "Analyze", "Reset", "Example", and "Help".

Please note that JavaScript™ is required.

# Auf Matrix basierende multilinguale Grammatikentwicklung

xmltext JTok TnT Chunkie Sprout SproutFS RASP PET result

```

TEXT the man smiles
TOP h1
RELS {
  prpstn_m_rel | _the_q |
  LBL h1 | ARG0 x9 pers=3 num=sg gender=n | h8 | _man_n |
  ARG0 h5 | RSTR h10 | ARG0 x9 pers=3 num=sg gender=n | h13 | _smile_v |
  BODY h12 | ARG1 x9 pers=3 num=sg gender=n | h15 |
}
HCONS {h5 qeq h15, h10 qeq h13}
ING {}

```

xmltext PET result

```

TEXT mannen smiler
TOP h1
RELS {
  mann-rel | def-q-rel | card_rel | smile-rel | prpstn_rel |
  LBL h3 | ARG0 x4 bounded=+ | h5 | LBL h8 | LBL h9 | LBL h1 |
  ARG0 x4 bounded=+ | RSTR h7 | ARG0 x4 bounded=+ | ARG0 e2 tense=present | ARG0 h10 |
  BODY h6 | CARG 1-rel | ARG1 x4 bounded=+ |
}
HCONS {h7 qeq h3, h10 qeq h9}
ING {}

```

# Auf Matrix basierende multilinguale Grammatikentwicklung

xmltext JTok TnT Chunkie Sprout SproutFS PET result

TEXT Der Mann lacht  
TOP h1

RELS	{	def	h3		_mann_n	h7		_lachen_v	h8		prpstn_m_rel	}		
		LBL											LBL	LBL
		ARG0 x6 pers=3 num=sg gender=m											ARG0 x6 pers=3 num=sg gender=m	ARG0 e2 tense=present
		RSTR											ARG0 x6 pers=3 num=sg gender=m	ARG1 x6 pers=3 num=sg gender=m
BODY	h5			ARG0	h9									

HCONS {h9 qeq h8, h4 qeq h7}

ING {}

xmltext ChaSen PET result

TEXT 人が笑う  
TOP h1

RELS	{	proposition_m_rel		_hito_n		_warau_v	}		
		LBL						LBL	LBL
		ARG0						ARG0	ARG0
		ARG0						RSTR	BODY
		h5		x7	h14	x7			

HCONS {h5 qeq h15, h14 qeq h6}

ING {}

xmltext JTok PET result

TEXT l uomo sorride  
TOP h1

RELS	{	_def_q		_uomo_n		_sorridere_v		proposition_m_rel	}		
		LBL								LBL	LBL
		ARG0								ARG0	ARG0
		RSTR								ARG0	ARG1
BODY	h7	x4	x4	h17	h17						

HCONS {h8 qeq h9, h17 qeq h1}

ING {}

# Jacy und Maschinelle Übersetzung

- Maschinelle Übersetzung gesprochener Sprache:  
Verbmobil (1992 – 2000)
- Dialogsprache
- Pragmatische Information
- Spontan gesprochene Sprache
- Nullpronomina, Ambiguitäten
- Unklare Satzgrenzen („Turns“ statt Sätzen)
- Fehler der Spracherkennung
- Kurze und fragmentarische Äußerungen
- Erkennung von Phrasen

# Jacy und E-Mail-Kategorisierung

- Sprache in E-Mails : oft kurz und fragmentarisch
- Spezielle Abkürzungen, Idiome und anderes Vokabular
- Tokenisierung, Integration von ChaSen
- Erweiterung des Lexikons
- Default-Einträge
- enge Kooperation mit HPSG-Grammatiken für andere Sprachen (Spanisch, Englisch) → Matrix
- Testumgebung für systematische Tests

# Jacy und Informationsextraktion

- **Projekt DeepThought:**
- Entwicklung von Grammatiken für mehrere Sprachen (Norwegisch, Spanisch, Deutsch, Italienisch)
- → Weiterentwicklung der Matrix
- Verknüpfung tiefer und flacher Verarbeitung
- → Entwicklung der „Heart of Gold“

# Jacy und Derivation ontologischer Information

- Basis: Definitionssätze aus Lexika

- z.B.:

ドライバー：自動車 を 運転 する人

doraib̄a: jidousha wo unten suru hito

driver: car acc drive do person

driver: a person who drives a car

- Relation: *driver is-a person*
- Extraktion von Hyponymie und Synonymie

# Jacy und Derivation ontologischer Information - 2

- Kurze und wohlstrukturierte Sätze
- breites Vokabular, viele Domänen
- → Erweiterung des Lexikons
- viele Nominalisierungen
- → Erweiterung der Grammatik
- Problem: Ambiguität in den Analysen
- → Treebanking und Aufbau eines Parse-Ranking-Modells

# Jacy und Maschinelle Übersetzung - neu

- **Jaen: Open-source Machine Translation:**
- Basis: Transfer über MRS-Strukturen
- kleine Menge von handgeschriebenen Transferregeln
- größere Menge von gelernten Transferregeln auf parallelen Daten
- Stochastische Verfahren für die Auswahl von Übersetzungen
- Vergleich mit statistischer MT (Moses):
  - Moses hat bessere BLEU und METEOR Scores
  - Jaen wird von Übersetzern besser bewertet

# Zusammenfassung

- JACY, Japanische HPSG:
  - Präzise syntaktische, semantische und pragmatische Information in Attribut-Wert-Strukturen
  - Kombination aus manuell erstellter lexikalischer Information und Default-Lexikoneinträgen
  - Multilingualer Kontext mit parallelen und konsistenten semantischen Ausgaben
  - Effiziente, hybride und robuste Verarbeitung
  - Eingesetzt in sehr unterschiedlichen Anwendungen

# Vielen Dank für Eure Aufmerksamkeit!

どうもありがとうございました。

```
TEXT  どうもありがとうございました。
TOP   h1
RELS  {
  | _doumoarigatougozaimashita_x | discourse_rel
  | LBL          h1                | LBL      h10001
  | ARG0         u3                | ARG0     h1
  | R-HNDL       h6                | R-HNDL   h6
}
HCONS {}
ING   {h1 ing h10001}
```