

長尾先生の言語観と例からの機械翻訳

辻井潤一

人工知能研究センター

国立研究開発法人 産業技術総合研究所

言語処理と言語学：AAMT前史の前史

- やまと：機械翻訳用計算機（和田弘：電総研）、1958年から
 - 計算をしない計算機、Categorial Grammar理論の先駆的工作、のちMUメンバー坂本義行さん

- ALPAC報告（1967年）機械翻訳に否定的、基礎研究とMATへのシフト

アメリカ構造主義言語学
L.ブルームフィールド
言語的相対論：E.サピア



- 1971年 辻井：京都大学修士課程 入学
- 坂井教授の研究室に所属
 - 長尾真先生は、助教授
 - フランス・グルノーブル大学（GETA、CNRS）から帰国
 - CKYアルゴリズム、意味を考慮した言語生成
 - 杉田繁治先生：機械翻訳
 - 同期：天野真家君（東芝入社、日本語ワープロ）
 - 河田勉さん（東芝から出向、仮名漢字変換の研究）、九州大学G
 - 金出武雄先生（博士課程、画像処理、CMUへ）
- 長尾先生から、言語処理の研究に誘われる



Syntactic Structure (1957)
Aspects of the Theory of Syntax (1965)
by N. Chomsky

構造主義 vs 認知主義

言語処理と人工知能：AAMT前史の前史

長尾先生の言葉

言語は人間の情報処理の根幹にある。言語処理技術は、将来の情報処理の中核になる。

Semantic Information Processing (ed:M. Minsky) 1968

Semantic Network: Ross Quillian

→A model of word perception called the Interactive Activation (IA) Model:
McClelland and Rumelhart (1981)

SHRDLU : T.Winograd (Ph.D Thesis, 1972) 積み木世界会話システム

LUNAR system (W.Woods, R.Kaplan, et.al. 1972) 月の岩石のQAシステム

→ Semantic Parsing

博士論文のテーマ：質問応答システムの研究 / IAx2

長尾先生に見られる二項対立

• 言語研究の方向

• 経験主義的指向性

➡ 西田龍雄（西夏文字解読、文学部）

- 記述言語学、構造主義言語学、コーパス言語学への傾斜、辞書学
- 統語論だけを取り上げるChomskyの立場に批判的

• 一方で、認知主義的な傾向も

➡ 梅本堯夫（認知心理学、教育学部）

- 意味や語用論、人間の内的処理、AIへの強い興味

• 研究への動機

井筒俊彦『意識と本質——精神的東洋を求めて』（1988）

• 社会的な使命感

- 工学の立場、研究成果の社会的な還元


➡ 機械翻訳プロジェクト（MU）

• 知的な探求心

- 人間の情報処理、人工知能

➡ 対話研究会
脳、理論言語学、認知心理学、AI
N. ChomskyやR.Montagueの論文

AAMT前史：80年代

- Coling 80 東京開催
 - 和田弘、長尾真、淵一博、石綿敏雄、坂本義行
 - M.Kay(Stanford), A.Zampoli(Piza), B.Vauquois (GETA),,,,,
 - 第5世代コンピュータプロジェクト (1982-1992)
 - 論理 (Prolog) によるAI (淵一博、電総研が中心、通産省)
 - MUプロジェクト (1982-1986)
 - 第2世代MT (科学技術論文の機械翻訳、長尾真、科学技術庁)
 - **工学に徹する、社会に貢献する技術**
 - 民間企業からの研究者・技術者 10名ほど (隅田さん、安達さんら)
 - EDR電子化辞書プロジェクト (1986-1994)
 - 民間会社8社 (共同出資、通産省)、横井俊夫、内田裕二。。。。
- 

百花繚乱の80年代機械翻訳

- 民間企業での機械翻訳開発の活発化
 - トランスファー方式、中間言語方式、理解に基づく翻訳
- アジア近隣諸国の機械翻訳プロジェクト（1987－1996）
 - CICC（国際情報化協力センター）、ODA
 - 中国、マレーシア、タイ、インドネシア、日本
 - 中間言語方式
- ATR自動翻訳電話研究所（1986－、名称変更あるが現在まで）
 - 音声翻訳
 - 第2世代方式、統計翻訳を経て、ニューラル翻訳へ
- Eurotraプロジェクト（1978－1992）
 - 野心的な第2世代、当初から迷走（在仏中の1981年当時でも）、言語学に強く傾斜
 - 辻井：1988年から英国マンチェスター大学でEurotra 6
- **MT Summit : Hakone、1987年9月（長尾眞）** → **AAMT, EAMT, AMTA, IAMT**

MT SummitからAAMT：社会的使命感

社会的な使命感

- **社会への貢献、機械翻訳を社会に広げていく**
- **多様なStakeholdersの合意形成**
 - **Decision Makers**：公的な政策決定者、民間企業の技術開発責任者
 - **Users**: 翻訳会社、翻訳者
 - **Technology Providers**: 技術者、研究者
- 評価の手法
 - 技術進歩の指標（主観評価：正確さ、流暢さ）→ Bleu値（定量評価）
 - Post-Editingのコスト、修正操作
- Translation Workflow全体のコスト
 - ドキュメント構成（章構成、図面）、ドキュメントのVersion管理
 - Translation Memory
 - 用語の統一的な管理システム(Terminology, 辞書学)

長尾先生の言語観：知的探求心

- **基本的な「ものの見方」**

- 既存研究の勉強はするが、とらわれず、対象と向き合う
- 壺のような芸術について、「良いものをたくさん見ることで、自分の鑑賞眼を養うこと」、ドグマ的ではない、Followerにはならない

- **言語観**

- 言語表現は様々な要因の重なりで発現している複雑なもの
- 無限集合を生成的に定義するChomskyの立場は、統語論のための理想化、工学には役に立たない → Chomskyはより内的なものへ純化、長尾先生は外的な言語現象へ
- 年齢、性差、社会クラス、Empathyの対象や話者の感情に応じて、言語が変化していくこと
- 膨大な有限集合
- 演繹的、無限を生成する規則の系よりも、類推や帰納、メタファーへなどへの関心
- コーパスの記述、辞書の構築などへの関心：構造主義的な傾向
- 同時に、人間の内的処理への関心：認知主義的な傾向

MUプロジェクト

- 基本的な立場

- アドホックな第1世代MTから系統的な方法論を持ったMT
- 変換方式
- 構造記述の階層性 形態素列→単語列→句構造→依存構造→意味構造
- 木構造間の写像、強力な構造変換規則
 - GRADE (中村順一)
 - PROLOG (第5世代) の始祖 (A.Colmerauer, グルノーブル)
- 再帰的変換：Compositional Translation (部分の翻訳→全体の翻訳)
- **無限を生成する規則の系でとらえる立場 (第2世代MTに共通)**
- 解析・変換過程におけるDisambiguation (周辺文脈、理解の関与)
- 言語は複雑な要因が絡み合ったもの、素性の束、技術的には未熟

合理主義からコーパス主義へ

- 例による翻訳（1984：長尾）
 - アカデミアのもの見方にとらわれず、問題と直接向き合う態度
 - 言語は膨大な有限集合：演繹的な体系より、類推やメタファー、帰納
 - 合理主義よりも経験主義、翻訳の認知的な機構への関心
 - 言語は多様な要因の相互作用
- 統計的翻訳（IBM, 1988年、90年、93年）
- Translation Memory（Sadler 1987年, Harris 1988年, Kay ?）
- Case-Based Reasoning（演繹よりも類推、1990年代に隆盛）
- Penn-Tree-Bank（1989 – 1996年）（EDR:1986-1994年）

おわりに：翻訳について、思うこと

MTの課題と人間翻訳者の役割

- テーブルの上のランプを消してください-> on かaboveか？
- 遊ぶ→to play, to hang around, to stay, to have fun, etc.
 - 一年間パリで遊んだ
 - 何人もの若者がコンビニで遊んでいる
 - 公園で子供たちが遊んでいる
- 意味と慣習
 - validate a ticket
- 言語使用者
 - 性別、年齢、職業、地域、年代：フライトでのサブタイトル
- 翻訳のむつかしさ
 - カフカの小説、Kunderaのコメント、法律名・新概念などの用語の翻訳