

# 言語の固有次元を測る

上田亮<sup>1</sup> 横井祥<sup>2,3</sup><sup>1</sup> 東京大学 <sup>2</sup> 東北大学 <sup>3</sup> 理化学研究所

ryoryoueda@is.s.u-tokyo.ac.jp yokoi@tohoku.ac.jp

## 概要

言語の複雑性に関する問題は言語学や計算機科学に属する複数のドメインで長い間探求されてきたが、近年では特に深層ニューラルネットワークモデルが獲得する自然言語文の表現の複雑性に注目が集まっている。これまでの表現学習モデルがどのような言語の複雑性を捉えてきたのか、またこれからのモデルがどのような複雑性を獲得していけばよいのかを考える契機となることを目的として、本稿では、固有次元の観点から言語の複雑性を推定する。実験の結果、最近のモデルから得られた表現であっても固有次元の推定値は数～十数程度となり、言語の持つ本質的な次元の小ささが示唆された。

## 1 はじめに

本稿の目的は、固有次元の観点から自然言語の複雑性について考えることである。固有次元 (Intrinsic Dimension) とは、大雑把に言えばデータセットの表現するのに必要最小限の変数 (軸) の個数のことである。表現学習によって推し進められた言語表現の複雑さを固有次元というレンズを通じて検討することで、十年來の表現学習のモデルの進化によっていかなる意味の軸が追加されてきたのか、大規模言語モデルの次元をどこまで上げれば良いか、といった困難な問題を解き明かすための一歩としたい。

言語の複雑性に関する探究は古くから存在する。20 世紀の統語理論や形式言語理論においては、オートマトンや文脈自由文法、文脈依存文法、Turing 機械などの形式的な道具立てを用いて言語をモデル化しようとする試みがなされ、Chomsky 階層という概念としてまとめられるに至った<sup>1)</sup>。また、計算言語学やコーパス言語学の方面においては、Zipf の法則

[4] や Heaps の法則 [5] などが知られている。これらの統計的普遍性質は共通して言語コーパスが大きければ大きいほど、そこに出現する語彙 (単語の種類数) も無際限に多くなっていくことを示唆しており、言語におけるスケールフリー性として知られている。また計算モデルの観点からの複雑性に関する議論も盛んである。言語データはマルコフ過程程度では到底モデル化できないような非常に長い距離の依存関係を有しているという指摘もある (e.g., 長相関) [6, 7]。Pitman-Yor 過程に代表されるような、言語を記号の生成過程としてモデル化し、言語と同等の複雑さをもつような生成過程を探求する向きもある [8, 9]。エージェント (ロボット) 間にゼロからコミュニケーションプロトコルを創発させる記号創発 [10] や言語創発 [11] も、言語現象を観察する代わりに構成 (construct) することを通して言語の複雑性に迫る試みである。さらには、Montague やその後続く論理学者たちが行ってきたような形式意味論的アプローチも、Frege の原理 (構成性原理) を前提として、文と意味の関係性に関する考察を重ねてきたという点において、言語の複雑性に関する一連の探究と見做せるかもしれない。

本稿では、言語のモデルおよびこれらのモデルが捉えるコーパスの複雑さを、固有次元の観点から明らかにすることを試みる。深層学習の利用が主流となっている近年の自然言語処理では、自然言語文  $x \in \mathcal{X}$  は往々にして何らかの (ニューラルネットワーク) モデル  $f_\theta: \mathcal{X} \rightarrow \mathcal{Z}$  を用いて表現 (representation)  $z \in \mathcal{Z}$  に変換される。このやり方によって多くの自然言語処理タスクが高い精度で解けていることから、深層学習モデルから得られる文の表現は、自然言語文の持つ様々な意味の情報を適切に保持していると考えられている<sup>2)</sup>。また通常、表現  $z$  は固定次元の実数ベクトル、すなわち連続的な

1) 特に近年では、文脈自由文法と文脈依存文法の間に位置する弱文脈依存文法が文の統語構造を良くモデル化できるという説が有力視され、組合せ範疇文法 [1, 2, 3] を代表とする弱文脈依存文法を用いた形式的・実験的な検証が 21 世紀の今日でも盛んに行われている。

2) 特に近年では大規模言語モデル (LLM) の性能向上が目覚ましく、もしかしたら (人類の知識や常識をも含めた) 言語の複雑性は高々数十～数百億個程度のパラメータに保存可能なのではないかとさえ思わせるほどである。

量として表されるため、離散シンボル列である生の自然言語文  $x$  よりもデータの性質を調べる上で扱いが容易である。以上のような理由から、本稿では、モデルから得られる表現と、表現の複雑さの度合いを自然に測れる固有次元推定という道具を通して、モデルとコーパスの複雑さの問題に迫ってみたい。

連続表現の集合の複雑さを表す量のひとつである固有次元の推定方法としては、近年高次元データ解析の文脈 [12] でしばしば用いられる **TwoNN** [13] を採用する。TwoNN はデータ点  $z, z'$  間の距離  $d(z, z')$  の情報を必要とするが、表現  $z$  は固定次元の実数ベクトルであるから、自然な距離として **Euclidean 距離**などを考えることができる<sup>3)</sup>。実験では、(ニューラルネットワーク) モデルとして Glove や BERT, GPT-2, SimCSE などの幅広いモデルを用いた。さらに、表現  $z$  を介さずに距離を定義できる Levenshtein 距離を用いた検証や、単純な単語の頻度情報を並べてベクトルとしただけの表現 (Bag-of-Words) をベースラインとして比較した。

実験の結果、今回用いたモデルやコーパスにおいては、自然言語文の推定固有次元は数〜十数程度であることが明らかになった。

## 2 TwoNN を用いた固有次元推定

大規模な高次元データは自然言語をはじめ自然界に幅広く存在するため、これを適切な方法で処理・分析するための方法が模索されてきた。高次元データの分析においては、しばしば多様体仮説が前提とされる。多様体仮説というのは、現実世界から得られる高次元データセットは、実は見かけよりも次元の小さな多様体上に分布しているはずだ、という考え方である。この上で、データセットが張る多様体の次元—固有次元 (Intrinsic Dimension; ID)—を推定することが、高次元データの分析における主要課題の一つとなっている。固有次元の推定はいくつかの難しさを伴う。たとえば、高次元データに潜む多様体は往々にして非 Euclidean であり、捩じれたり曲がったりしているものである。また、実際に得られるデータセットが潜在的な多様体の上に均一に分布しているとは限らない。

これに対し、[13] は **TwoNN** という新たな固有次元推定アルゴリズムを提案し、上に挙げたような問題がある程度克服できることを理論・実験の両面か

ら示した。

興味を持たれる読者のために、TwoNN の直感的な計算手順と利点を詳述する。TwoNN はデータ  $z, z'$  間の距離  $d(z, z')$  の情報を用いて固有次元を推定する。各データ  $z$  について、距離関数  $d$  に関して  $z$  に最も近い別のデータ  $z^{(1st)}$ , 2 番目に近いデータ  $z^{(2nd)}$ , 3 番目に近いデータ  $z^{(3rd)}$ , ……と順にデータを並べていく状況を考えてみよう。このとき、 $z$  が属している多様体の次元が大きいほど  $z^{(1st)}, z^{(2nd)}, z^{(3rd)}, \dots$  の系列が  $z$  から“遠ざかっていくスピード”も大きくなっていくはずである。特に、潜在的な多様体が全体的には捩じれたり曲がっていたりしたとしても、最初の 2 点  $z^{(1st)}, z^{(2nd)}$  は  $z$  にごくごく“近い”場所にあるはずなので、多様体の複雑な構造の影響をそれほど受けることなく  $z$  の周りで“均一に”分布しているデータのサンプルと思うことができ、それ故により正確に“遠ざかっていくスピード”を推定できそうである。このことを理論面から丁寧に説明し、シンプルなアルゴリズムとしてコンパクトにまとめたことが [13] の主要な貢献であり、TwoNN の利点である。

TwoNN の具体的なアルゴリズムは以下の通り。データセット  $\{z^{(i)}\}_{i=1}^N$  および距離関数  $d: \mathcal{Z} \times \mathcal{Z} \rightarrow \mathbb{R}_{\geq 0}$  を入力として、

1. 各点間の距離  $\{d(z^{(i)}, z^{(j)})\}_{1 \leq i < j \leq N}$  を測る。
2. 各  $1 \leq i \leq N$  について、計算した距離に基づき  $z^{(i)}$  に最も近い 2 つの点を見つけ、その距離をそれぞれ  $r^{(1st)}, r^{(2nd)}$  とする。
3. 各  $1 \leq i \leq N$  について  $\mu^{(i)} = r^{(2nd)} / r^{(1st)}$  とおく。
4.  $(\mu^{(1)}, \dots, \mu^{(N)})$  を昇順に並べ替えた結果得られる順列を  $\sigma: \{1, \dots, N\} \rightarrow \{1, \dots, N\}$  とし、各  $1 \leq i \leq N$  について  $F^{\text{emp}}(\mu^{(\sigma(i))}) = i/N$  とおく。
5.  $\{(\log \mu^{(i)}, -\log(1 - F^{\text{emp}}(\mu^{(i)})))\}_{i=1}^N$  という形で表されるデータセットを考え、これに当てはまるように直線回帰をする。回帰の結果得られる傾きが固有次元の推定値である。

## 3 問題設定と実験設定

### 距離の取り方について

特に今回は「文」を分析の単位とし、文間の距離を定めた上で TwoNN により固有次元を推定する。すなわち、距離が入っている空間（ないしそれを計算可能にする各文埋め込み空間）のそれぞれで、文

3) Levenshtein 距離のような、表現  $z$  を介さずに直接文間の距離  $d(x, x')$  を定めることのできる関数を用いてもよい。

の集合がどの程度の固有次元（広がり）を持つのかを推定する．具体的な距離構造・表現モデルとして以下を扱う：

- 文字を単位とする Levenshtein 距離 (編集距離).
- 単語を one-hot 符号化しておき，文に出現する単語の one-hot 符号を足し合わせて得られるベクトルを文の表現 (Bag-of-Words; **BoW**) としたときの，表現間の Euclidean 距離.
- 単語を Word2Vec 系モデルで符号化しておき，文に出現する単語の符号の平均をとって得られるベクトルを文の表現 (所謂分散表現) としたときの，表現間の Euclidean 距離.
- BERT 系 Transformer モデルに文を入力したときに得られる，[CLS] の埋め込み表現を表現としたときの，Euclidean 距離.
- GPT 系 Transformer モデルに文を入力したときに得られる，最後のトークンの最終層の隠れ状態を表現としたときの，Euclidean 距離.

Word2Vec 系のモデルとしては，GenSim ライブラリ [14] から得られる ‘glove-wiki-gigaword-50’ (以降 **Glove-50** とする) 及び ‘glove-wiki-gigaword-300’ (**Glove-300**) [15] を用いた．BERT 系のモデルとしては，HuggingFace ライブラリ [16] から得られる ‘bert-base-uncased’ (**BERT**) [17] 及び対照学習モデルである ‘princeton-nlp/unsup-simcse-bert-base-uncased’ (**SimSCE**) [18] を用いた．GPT 系のモデルとしては，HuggingFace ライブラリから得られる ‘gpt2’ (**GPT-2**) [19] を用いた．

## 取り扱うコーパスについて

また，今回対象とする自然言語は，NLTK ライブラリから簡単にダウンロードできる幾つかの英語のデータセットとする．具体的には *The Reuters-21578 benchmark corpus*, *ApteMod version* (reuters) と *Web Text Corpus* (webtext) を採用する．これらのデータを文ごとに区切り，小文字に統一し，改行文字を取り除き，連続するスペース文字列を 1 文字に置き換えた．

今回採用するコーパスには，それぞれ数万文が収録されている．各コーパスに含まれるすべての文ペアについて距離を計算するのは計算コストがかかるので，今回はコーパスから 1,024 文ずつ非復元サンプリングをしてミニバッチとし，ミニバッチ内で固有次元を推定したのち，それらの平均と標準誤差を

掲載するという形をとることとする．勿論十分な時間と計算リソースをかければ全ての文ペアについての距離を計算することは可能である．これについては今後の課題とする．

## 4 実験結果

	reuters	webtext
Levenshtein	4.14 ( $\pm 0.04$ )	6.75 ( $\pm 0.13$ )
BoW	9.81 ( $\pm 0.08$ )	15.19 ( $\pm 0.34$ )
Glove-50	7.88 ( $\pm 0.04$ )	10.01 ( $\pm 0.18$ )
Glove-300	8.07 ( $\pm 0.05$ )	11.60 ( $\pm 0.27$ )
BERT	12.10 ( $\pm 0.04$ )	12.27 ( $\pm 0.07$ )
SimCSE	9.71 ( $\pm 0.05$ )	9.18 ( $\pm 0.09$ )
GPT-2	7.27 ( $\pm 0.02$ )	8.92 ( $\pm 0.03$ )

表 1 TwoNN を用いた固有次元の推定結果．小数点第 3 位を切り捨てて表示している．( $\pm$ ) は標準誤差を表す (こちらも小数点第 3 位切り捨て)．

TwoNN を用いて自然言語文の固有次元を推定した結果を表 1 に示す．少なくとも今回の設定においては，推定固有次元は数～十数次元程度に収まるようである．ただし，今回対象としている文の表現 (representation) がある種の不可逆圧縮になっている可能性に注意する必要がある．BoW は言語の語順に関する情報を捨てているし，Word2Vec (Glove-50, Glove-300) はそこからさらに one-hot 符号化された単語を embedding として圧縮していることになる．BERT や SimCSE の [CLS] トークンはあくまでテキスト分類に役立つ情報を保持しているにすぎず，文の情報全てを保存しているとは限らない．GPT-2 に関しては，最後のトークンに対応する隠れ状態のみを用いているため，せいぜい次のトークンを予測するのに役立つ程度の情報しか保持されていないかもしれない．つまり，仮に文の情報全てを保持する表現があったとするならば，その表現の固有次元は今回得られた結果よりも大きい可能性がある．

個別の結果を見ていくと，まず Glove-50/300 の推定固有次元が，BoW のそれより小さいことに気が付く．reuters に関して言えばその差はわずかである．これは，one-hot 符号化された単語を embedding として大幅に圧縮していることに起因すると考えられる．また，BoW の推定固有次元がせいぜい数～十数程度しかないことを踏まえれば，one-hot 符号ではなく，高々数十次元程度の embedding として表現する Word2Vec の方法論の合理性が伺える．また，BERT と SimCSE の結果を比べると SimCSE の



推定固有次元のほうが小さくなっている。これは、対照学習によって表現が単純化されたことに起因すると考えられる。テキスト分類のタスクにおいて、しばしば事前学習済みの Transformer モデルに線形分類器を取り付けたものが用いられることを踏まえれば、こうした単純化は(線形写像にも分離可能な単純な空間を作るという意味において)合理的に思われる。なお、GPT-2 の推定固有次元は Word2Vec のそれよりも小さいが、これは上でも述べたように、次のトークンを予測するのに役立つ程度の情報しか保持していない可能性が考えられる。

## 5 構造を保つ写像としての言語：言語創発の観点から

本節では、言語創発の分野と関連付けながら、言語の複雑性に関するさらなる考察を綴る。人工エージェントの間でコミュニケーションをさせ、言語の創発をシミュレーションを試みる言語創発(創発コミュニケーション; EC)は、構成論的なアプローチによって言語の複雑性に迫ろうという挑戦的な分野である。人工エージェント間に生じるコミュニケーションプロトコル—創発言語—is、しばしば(“話す”行為に相当する)意味  $z \in \mathcal{Z}$  から記号  $x \in \mathcal{X}$  への写像  $S: \mathcal{Z} \rightarrow \mathcal{X}$  もしくは(“聞く”行為に相当する)記号から意味への写像  $R: \mathcal{X} \rightarrow \mathcal{Z}$  として定義される。創発言語は我々人間の話す自然言語とは当然異なるので、何らかの方法で創発言語の「自然言語らしさ」を評価する必要があるのだが、言語創発でよく用いられるのは Topographic Similarity (TopSim) [20, 21] という評価指標である。大雑把に言えば、TopSim は写像  $S$  の等長写像的な振舞いを評価する指標である。即ち、ある固定の  $k \in \mathbb{R}_{>0}$  が存在して、任意の  $z, z' \in \mathcal{Z}$  について  $d_{\mathcal{Z}}(z, z') = k d_{\mathcal{X}}(S(z), S(z'))$  となると、TopSim は最大値になる<sup>4)</sup>。  $d_{\mathcal{Z}}, d_{\mathcal{X}}$  はそれぞれ  $\mathcal{Z}, \mathcal{X}$  上の距離関数である。TwoNN の理論とアルゴリズムが示唆しているように、データ間の距離はデータセットの固有次元の情報を含んでいるわけであるから、TopSim は暗に「意味空間と記号空間の(固有次元的な意味での)複雑性を同等にせよ」あるいは「写像  $S: \mathcal{Z} \rightarrow \mathcal{X}$  は距離構造を保存する写像であれ」という要請を評価の基準にしていることになる。興味深いことに、EC の研究者は TopSim のことを構成性 (Compositionality) の評価指標だと思っている。関数の等長写像性という観点から TopSim を

紹介された後に、それが構成性の指標であるといわれるとなんだか不思議なことに思われるかもしれない。しかし、EC における「構成性」は、Type Logical Grammar (TLG) [22] のような、形式的に構成性を取り扱う言語理論分野における「構成性」と案外よく似た考え方をしているのではないかと考えられる。TLG においては、文や意味を群に似たものとして定義し、構成的な semantic parser を群準同型に似たような形で定義する<sup>5)</sup>。つまり、TLG においては(距離ではなく)何らかの演算によって作られた構造を保存するような写像が構成的な言語であると考えられているわけである。従って、EC における「構成性」も TLG における理論言語学的な「構成性」も、「何らかの構造を保つように写像すること」という点で価値観を共有している。

Frege の原理 (構成性原理) がある程度正しい想定であると信じる場合、“聞く”行為や“話す”行為は単に「(ある程度) 構造を保存する写像」であるということになる。写像としての言語に何らの制約も課さないとすると議論のしようがなくなってしまうので、こういった制約を課すこと自体は至極当然のことと思える。ただし、そうだとすると、本稿で我々が言ってきたような言語の複雑性や固有次元といったものの正体は「意味」にあって、単に言語は「意味」の複雑性を投影する「記号」にすぎないということなのだろうか？ 逆に、言語にまつわる統計的必然性に関する研究が主張しているように、「記号」こそが複雑性の正体であって、それを「意味」に投影するから人間は複雑な意味を操作できるようになったのであろうか？ あるいは、重要なのは「意味」と「記号」の相互作用であって、どちらか一方だけで十分な複雑性が生じることはないのであろうか？ 恐らくどれも作業仮説としては妥当な問いである。重要なのは、今後も様々な問いを通して言語の複雑性を探求していくことであろう。

## 6 まとめと今後の展望

本稿では、言語の複雑性を測ることを目的として、自然言語文の固有次元を TwoNN を用いて推定した。結果として、今回の設定では自然言語文の推定固有次元は数～十数程度となった。より多くの言語やコーパス、モデルを対象とする調査は今後の課題である。

4) 実際には、この等式から多少逸脱していても良いような、もう少し“融通の利く”評価指標ではある。

5) 正確には群とは限らない。例えば、非結合的な言語 NL や NLP など。この説明はあくまで直感的なものである。

## 謝辞

本研究は JSPS 科研費 JP23KJ0768, JP22H05106 の助成を受けたものです。

## 参考文献

- [1] Mark Steedman. **Surface structure and interpretation**, Vol. 30 of **Linguistic inquiry**. MIT Press, 1996.
- [2] Mark Steedman. **The syntactic process**. Language, speech, and communication. MIT Press, 2000.
- [3] 戸次大介. 日本語文法の形式理論-活用体系・統語構造・意味合成-. 日本語研究叢書 (フロンティアシリーズ). くろしお出版, 2010.
- [4] George K. Zipf. **Human Behaviour and the Principle of Least Effort**. Addison-Wesley, 1949.
- [5] H. S. Heaps. **Information Retrieval: Computational and Theoretical Aspects**. Academic Press, Inc., 1978.
- [6] Henry W. Lin and Max Tegmark. Critical behavior in physics and probabilistic formal languages. **Entropy**, Vol. 19, No. 7, 2017.
- [7] 田中久美子. 言語とフラクタル. 東京大学出版会, 2021.
- [8] Y. W. Teh. A Bayesian interpretation of interpolated Kneser-Ney. Technical Report TRA2/06, School of Computing, National University of Singapore, 2006.
- [9] Yee Whye Teh. A hierarchical Bayesian language model based on Pitman-Yor processes. In Nicoletta Calzolari, Claire Cardie, and Pierre Isabelle, editors, **Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics**, pp. 985–992, Sydney, Australia, July 2006. Association for Computational Linguistics.
- [10] Tadahiro Taniguchi, Yuto Yoshida, Akira Taniguchi, and Yoshinobu Hagiwara. Emergent communication through metropolis-hastings naming game with deep generative models, 2023.
- [11] Angeliki Lazaridou and Marco Baroni. Emergent multi-agent communication in the deep learning era, 2020.
- [12] Alessio Ansuini, Alessandro Laio, Jakob H. Macke, and Davide Zoccolan. Intrinsic dimension of data representations in deep neural networks. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d’Alché-Buc, Emily B. Fox, and Roman Garnett, editors, **Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada**, pp. 6109–6119, 2019.
- [13] Elena Facco, Mariad’ Errico, Alex Rodriguez, Alessandro Laio. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. **Scientific reports**, Vol. 7, No. 1, p. 12140, 2017.
- [14] Radim Rehurek and Petr Sojka. Gensim–python framework for vector space modelling. **NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic**, Vol. 3, No. 2, 2011.
- [15] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics.
- [16] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Huggingface’s transformers: State-of-the-art natural language processing, 2020.
- [17] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Tamar Solorio, editors, **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [18] Tianyu Gao, Xingcheng Yao, and Danqi Chen. SimCSE: Simple contrastive learning of sentence embeddings. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 6894–6910, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [19] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.
- [20] Henry Brighton and Simon Kirby. Understanding linguistic evolution by visualizing the emergence of topographic mappings. **Artif. Life**, Vol. 12, No. 2, pp. 229–242, 2006.
- [21] Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark. Emergence of linguistic communication from referential games with symbolic and pixel input. In **6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings**. OpenReview.net, 2018.
- [22] Michael Moortgat. Chapter 2 - categorial type logics. In Johan van Benthem and Alice ter Meulen, editors, **Handbook of Logic and Language**, pp. 93–177. North-Holland, Amsterdam, 1997.