

目的：対話者の関係と談話の目的の異なる5つのコーパスを分析し、それらの間に見られる語彙カバー率の差異とそれぞれのコンテキストにおける談話遂行可能なレベルに必要な語彙数を調査する。

背景

- ① 談話に対する語彙のカバー率 (2000WF で95%~99%の幅が見られる)
Schonell et al. (1956) オーストラリアの労働者間の談話を調査し、2000WF が99%をカバーしている …コーパスデータの種類・サイズ(512,000)に限界
Adolphs & Schmitt (2003) 5百万語のコーパスを用い、2000WF では談話の95%のカバー率とした … 発話の種類を範疇分けせず、簡略化しすぎ ↔ コンテキストの異なる発話には、少なくともある程度異なる性格を持っているため
- ② 発話の「コンテキスト」と「特徴」(談話のコンテキストによる特徴的な発話)
Stenström (1990) 異なる発話コンテキスト下におけるディスコースマーカの頻度と機能を調査し、インタラクティブ度とディスコースマーカの頻度の相互作用
McCarthy (1998) コンテキストと直示語・語彙の関わりに **Stenström (1990)**同様の傾向
Carter & McCarthy (1995) 談話のタイプの違いと文法的特徴の関係、談話と談話の設定状況(例：動作中の言葉)の関係
- ③ 談話へ影響を与えるコンテキスト：談話はコンテキストに影響を受けている
Halliday & Hasan (1985) フィールド(談話の環境)・テナー(談話の当事者とその関係)・モード(コンテキスト下での言語の役割)が談話の本質を形成する

しかし、口語の談話が異なるコンテキストでどのように異なるかに関する研究は比較的少なかった。結果的に口語の様々なコンテキスト下での語彙カバー率に焦点が当てられてこなかった。(口語のコーパス・適切な範疇分けを持つコーパスの欠如に起因)

CANCODE (Cambridge and Nottingham Corpus of Discourse in English)

1994~1999 に渡り主なデータ収集が行われた。年齢・性別・素姓・教育レベルの異なる成人話者の様々な談話環境・ジャンルにおける会話を収集

5つのコンテキストによる範疇分け

- | | |
|---------|---|
| private | 1. intimate … 話者同士の距離が最短。話者がこの範疇だと認めた場合のみ |
| ↑↓ | 2. socio-cultural … 会話のみを目的とした場合 |
| | 3. professional … 仕事の一部としてのやりとり |
| public | 4. transactional … 知らぬ者同士が取引を達成する目的 |
| | 5. pedagogic … 指導・セミナー講義等 = アカデミック/トレーニング |

方法

* ワードカウント：ワードファミリーではなく、1語形を1語としてカウント

1. CANCODE よりコンテキストごとの頻度リストの作成
2. 2000~5000 の頻度帯までにある語数を総語数で割り、カバー率を算出
3. 総語数の違いによる影響を検証 → 結果的に影響はない

結果と考察

● 表 2 より、コンテキストの違いにより、語彙カバー率に顕著な違いが見られた。(カイ2乗検定では有意差なし)

● 最高カバー率 (Transaction) と最低カバー率 (pedagogic) の間には、5000 語レベルで、1.7 ・ 2000 語レベルで 3.9 ポイントの差が見られた

● Pedagogic のコンテキストで、Transaction と同等のカバー率を得るには、2000 語レベルで 3,608 語、5,000 語レベルでは、7,307 語必要であることが明らかになった

Table 2. Percentage of lexical coverage of five speech genre categories

Category (Total tokens in sub-corpus)	2,000 word forms % of coverage	3,000 word forms % of coverage	4,000 word forms % of coverage	5,000 word forms % of coverage
<i>Pedagogic</i> (456,177)	90.36	93.15	94.90	96.11
<i>Intimate</i> (957,192)	92.81	94.70	95.84	96.63
<i>Socio-cultural</i> (1,709,598)	92.43	94.34	95.51	96.31
<i>Professional</i> (480,627)	93.28	95.28	96.51	97.35
<i>Transactional</i> (1,166,825)	94.30	96.10	97.14	97.82

傾向として、**private/interaction** から **public/transaction** へと変化するに従って語彙のカバー率は高くなるが、当事者同士の親密度はそれほど影響しておらず、目的達成のための談話 (Professional と Transitional) になると急激にカバー率は上昇する。→ トピックの範囲の広さが影響していると推測される。

Pedagogic のコンテキストはカバー率がもっとも低い結果となった。

- (1) アカデミック/トレーニングの談話に対処するには、日常会話よりも広い語彙が必要であることの表れ
- (2) EAP において語彙の要素を含めるための議論材料となる

すべての頻度帯において、**Transactional > Professional > Intimate > Socio-cultural > Pedagogic** という順番が成り立っており、これはすべての頻度帯にコンテキストの影響があることを示唆している。

限界

ワードファミリー換算での調査の必要性

Discussion Question

1. Private な会話よりも Public な会話の方がカバー率が高いという結果は直感に反している様に思われるが、なぜこのような現象が起こるのか。