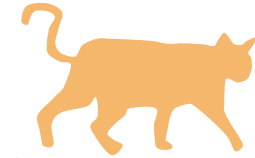


談話情報を用いた日本人学習者による 論説文の計量分析

小林雄一郎
(大阪大学)

第4回NLP若手の会シンポジウム
(2009年10月1日、京都大学)

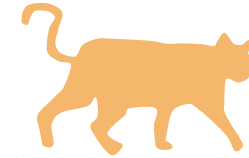
はじめに



- × 近年の英語教育の現場における実践的コミュニケーション能力の育成
 - × 「何を」伝えるかよりも、「いかに」伝えるかが必要不可欠
 - × 広義での談話標識の重要性
- ↓
- × だが、対象言語で適切に談話標識を用いることは難しい



本研究の目的

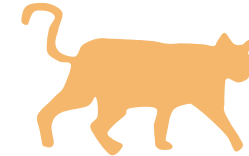


- × 談話標識の使用に関して。。。
 - 日本人中学生、高校生、大学生の学習者コーパスに関する縦断的研究
 - 日本人学習者と母語話者の差異に関する分析
 - 日本人学習者の特徴抽出と誤用分析

- × そのために。。。
 - 談話標識の自動タグ付けプログラム
 - テキストマイニング(データ縮約、テキスト分類)



コーパス

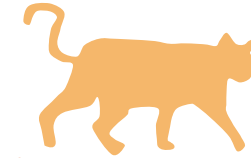


- × JEFLL Corpus (日本人中学生、高校生)
- × ICLE-JP (日本人大学生)
- × LOCNESS (アメリカ人大学生)

	JH	SH	UNI	NS
Corpus	JEFLL		ICLE-JP	LOCNESS
Tokens	162919	179750	168800	149574

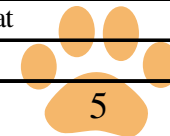


談話標識の定義

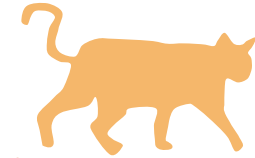


- ✕ HylandのMetadiscourse Markersのリストに基づく
 - 10種類の意味カテゴリーに属する約400種類の談話表現を網羅的にリスト化
 - これまでも、様々な計量分析に応用

Category	Function	Examples
Interactive resources	Help to guide reader through the text	
Transitions (TRA)	Express semantic relation between main clauses	in addition / but / thus / and
Frame markers (FRM)	Refer to discourse acts, sequences, or text stages	finally / to conclude / my purpose here is to
Endophoric markers (END)	Refer to information in other parts of the text	notes above / see Fig / in section 2
Evidentials (EVI)	Refer to source of information from other texts	according to X / (Y, 1990) / Z states
Code glosses (COD)	Help readers grasp functions of ideational material	namely / e.g. / such as / in other words
Interactional resources	Involve the reader in the argument	
Hedges (HED)	Without writer's full commitment to proposition	might / perhaps / possible / about
Boosters (BOO)	Emphasize force or writer's certainty in proposition	in fact / definitely / it is clear that
Attitude markers (ATM)	Express writer's attitude to proposition	unfortunately / I agree / surprisingly
Engagement markers (ENG)	Explicitly refer to or build relationship with reader	consider / note that / you can see that
Self-mentions (SEM)	Explicit reference to author(s)	I / we / my / our



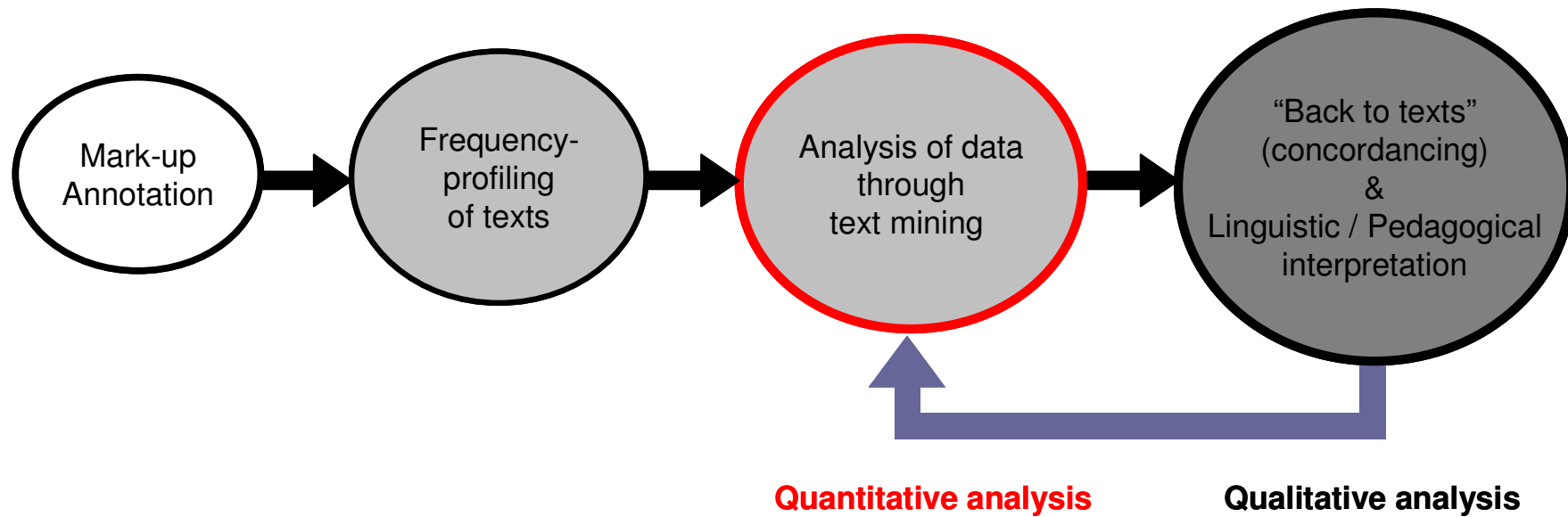
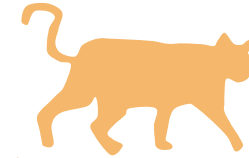
Research questions



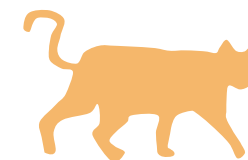
- × 中学生、高校生、大学生と学年が上がるにつれて、談話標識の頻度や使用傾向はどのように変化するのか
- × 学習者(中学生、高校生、大学生)と母語話者の間には、談話標識の頻度や使用傾向にどのような違いがあるのか



方法論



頻度集計

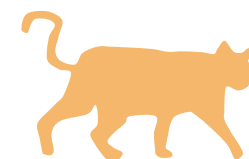


× 粗頻度

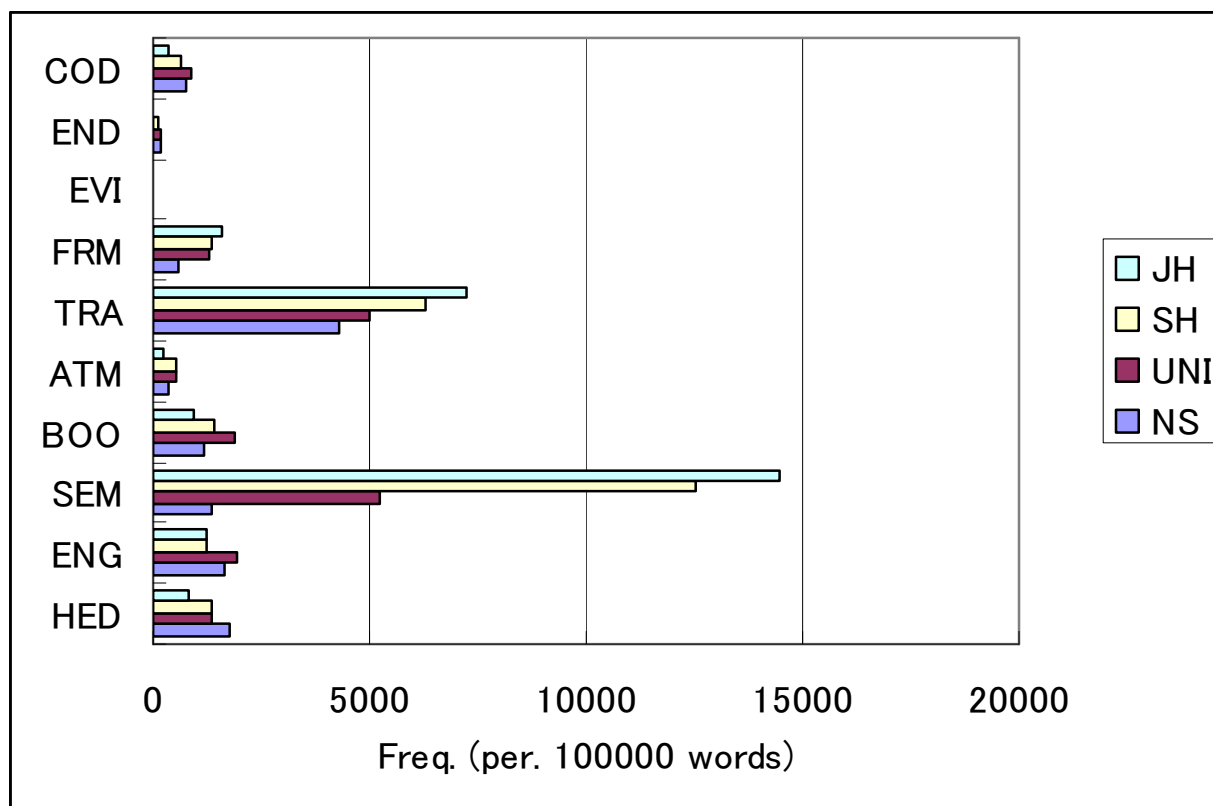
	JH	SH	UNI	NS	chi-squared	<i>p</i> -value
COD	568	1133	1475	1175	1480.964	***
END	114	195	334	306	496.623	***
EVI	0	4	44	82	1685.416	***
FRM	2573	2442	2213	853	162.497	***
TRA	11788	11325	8443	6455	741.850	***
ATM	419	985	887	503	419.840	***
BOO	1550	2506	3170	1766	1735.963	***
SEM	23622	22549	8869	2018	12992.040	***
ENG	1985	2261	3242	2470	2422.122	***
HED	1343	2385	2291	2645	3037.304	***

df =3, *** = $p < .001$





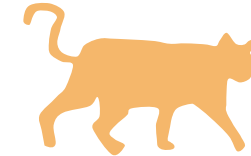
× 10万語あたりの相対頻度



Yuichiro KOBAYASHI @ YANS2009

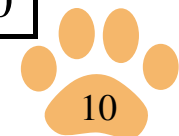


相関分析

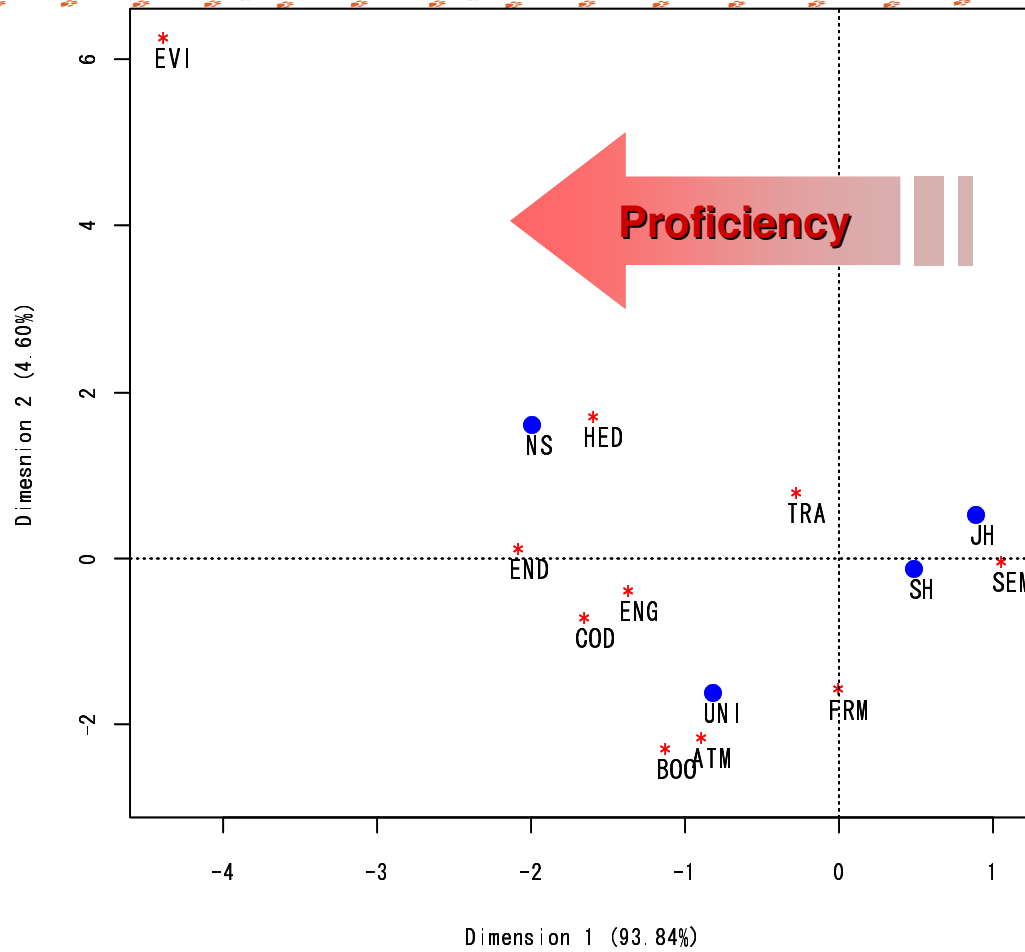
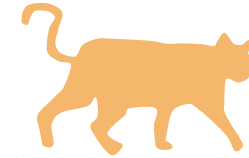


- × 相関分析学習者間では相関係数は0.910～0.998と極めて高い
- × 母語話者と学習者では0.455～0.760と相対的に低い
- × 学習者は、徐々に母語話者に近づいていく

	JH	SH	UNI	NS
JH	1.000			
SH	0.998	1.000		
UNI	0.910	0.916	1.000	
NS	0.455	0.466	0.760	1.000



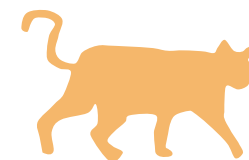
対応分析



Yuichiro KOBAYASHI @ YANS2009

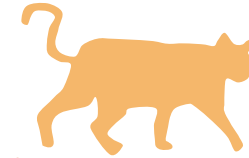


Case study: Transitions (TRA)

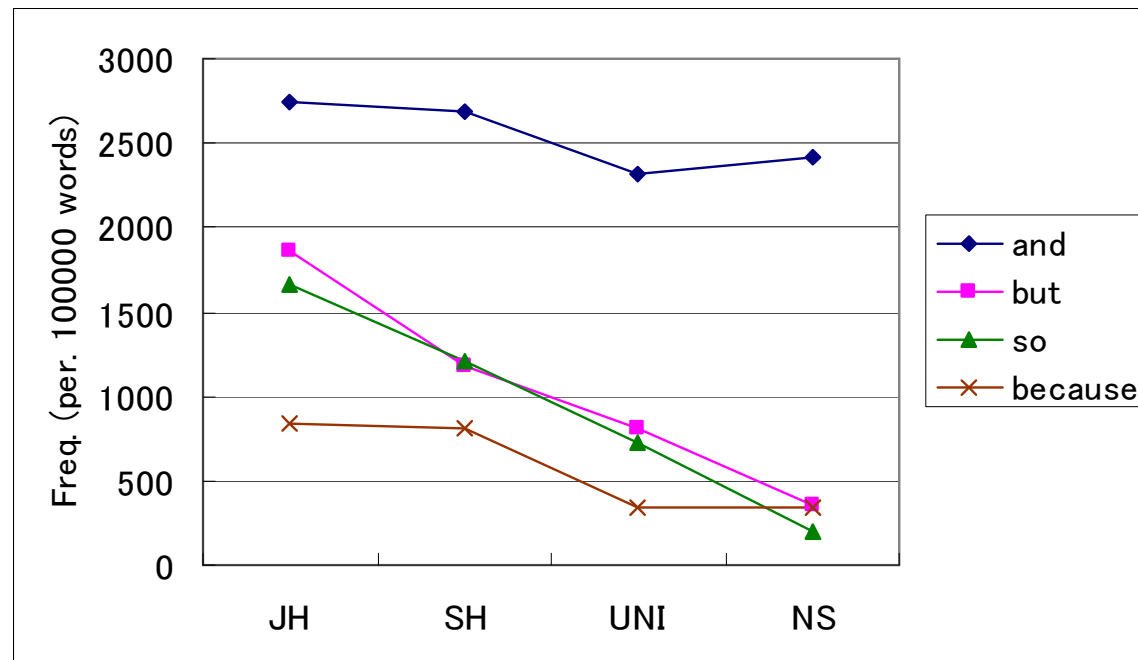


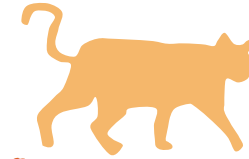
- × 学習者による産出言語の中核を担うのは内容語であるが、機能語も文構成上で重要な働き
- × 接続詞は、初期段階から英語学習教材などで多く提示され、学習者に馴染みのある語彙
- × だが、接続詞を適切に使用することは、語彙に関する知識のみならず、構文・意味・談話に関する知識が要求されるため、必ずしも容易ではない



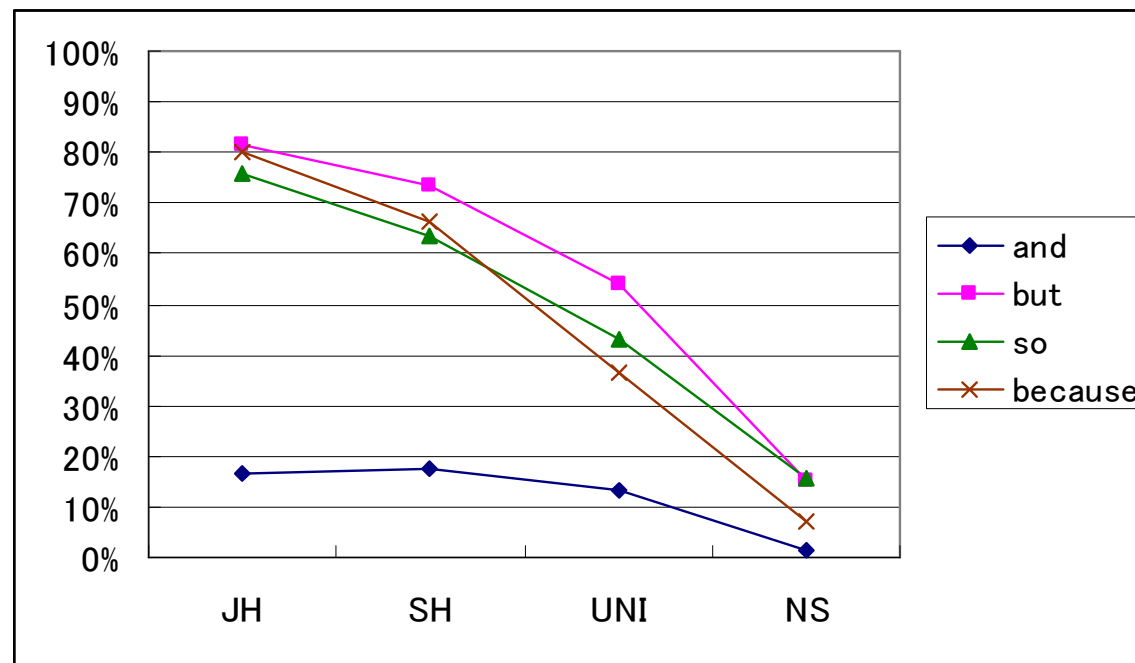


- 分析データにおけるTRAの頻度のうち、90.03%は頻度上位4語 (*and*, *but*, *so*, *because*) によるもの



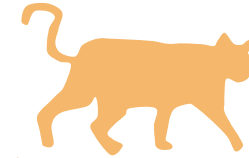


■ 頻度上位4語が文頭で生起する割合



■ 書き言葉では、文頭の*and*、*but*はまれ

典型的な誤用例



× 従属節を持たない断片文 (*so, because*)

■ * So, I want it very much.

■ * Because it is my dream.

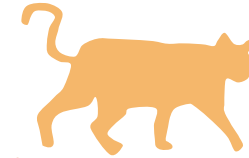


■ 日本語(「だから」、「なので」)の影響

■ 話し言葉の影響 etc.

■ 特に、*because*の場合、中学生、高校生、大学生が文頭で使ったとき、それぞれ97.55%, 94.75%, 84.58%は断片文





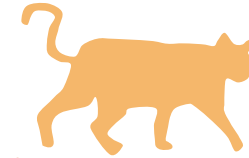
× unnecessary comma insertion

- * And, I like books and games.
- * But, it's very interesting.
- * So, I have only four thousand yen now.
- * Because, we need money.



- とりわけ、*and*と*but*に多い
- コンマが必要な接続副詞 (e.g. *also*, *however*) との混同か





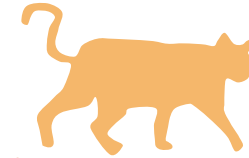
× 相関接続詞に関する誤り

- * I don't have tennis shoes and a hat!
- * I don't have rice and miso-soup in the morning.
- * Though, we can't still speak and write English fluently.



- 否定語のあとでは、*or*ないしは*nor*が正しい





× 接続詞の二重使用

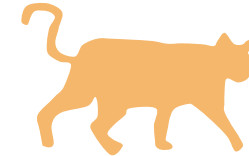
- * Although he died soon, but his story, and his name as a story teller, lasted forever.

× 接続詞の多重使用

- * If I take all then I will die because there are very heavy, so I can't escape.
- 他の誤用例や質的分析に関しては、予稿を参照

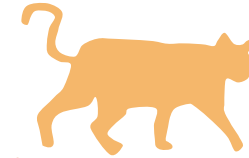


教育的示唆



- 接続詞の生起位置や不要な句読点の挿入のような誤用に対しては、DDL (Data-driven learning) 的な発想で、母語話者の使用例をKWIC (Key Word In Context) concordanceで示すことが有効
- 断片文のような構文の誤用に対しては、前の文章と接続させる指導や、文頭の*because*を*That is because*のような表現に置き換えるような指導がもっと必要





- 接続詞の過剰使用は、学習者が自らの文章の結束性に自信がないことの表れであり、それに対して、代名詞のような他の言語表現も結束性を持つことを提示し、文や節をつなぐ際に接続詞以外の選択肢を学習者に与えることも必要
- 教科書の「論理的な文章を書くための便利な表現」のリストがあるが、それらの表現を「多く」使うことが必ずしも「効果的に」使うことを意味しないということにも注意を喚起

