

多様な様式を網羅した会話コーパスの共有化

リーダー： 伝 康晴（言語資源研究系：客員教授）

日常の会話は様々な様式



我々の日常的な談話行動を理解するには、様々な様式を網羅した会話コーパスが必要

But

新規収録・転記は負担高



既存の会話コーパスの共有化

問題点：転記基準の不統一
基本情報の欠如

共有化の問題点を解決

1. 転記基準の共通化
2. 基本情報の仕様策定
3. 基本情報の付与
4. 共有コーパスの基礎的分析
5. コーパス共有の有効性確認
6. 適用範囲の調査

0017 00024.966-00027.194 R:
うち & ウチ
娘は & イモトワ
遅くね & スゴクネ
(F あの一) & (F アン一)
0018 00028.253-00029.388 R:
海外 & カイガイ
好きで & スキデロ
0019 00029.425-00031.454 L:
(F へー) (F ええ) (F ええ) & (F へー) (F エー) (F エー)
0020 00029.851-00031.738 R:
(F うー) & (F ウー)
行ってるんですけど & イッテルンデスケド
私は & ワタシワ
0021 00031.958-00035.409 R:
結局 & ケツキョク
後にも & アトニモ
先にも & サキニモ
あの & あの
中国にしか & チューゴク(ワ ディ:ニ)シカ
行ってなくて & イッテナク(の)テ
0022 00035.124-00035.269 L:
(F あ) & (F あ)
0023 00035.525-00035.319 L:
そうなんだ & ソーナンダ

01C: なあ、話し合【いで決めたらええねん
02B: 【内緒でええなあ
03C: なあ
04B: 世界が(,)平和になったらいいと思うし
05C: いいのって思うよ
06: ()
07B: う:ん
08: ()
09A: 世界が平和か
10B: ね
11A: 久しぶりでした
12B: うーん
13C: まますな:自分が幸せになって【から
14B: 【そうや
15C: =世界のこと考えるけど=
16B: =うち幸せちゃうからな
17C: まず自分が【幸せになったら世界の幸せ=
18A: 【あら
19C: =をな

行	J1	F	J2
1			
2			あ、さっきしゃべってたの
3			ってなんだったんですか
4	あ、あれは旅行に行っただ		なんかドイツとかー
5		あーそうですねーあの一	
6		お姉さんと一緒に	
7			ドイツ↑
8		ドイツとフランス、行ってきた	
9		2月かな！ちょっと寒いー	
10		寒かった	

大規模会話コーパスの開発プロジェクトにつなげる