

# 会話分析方式への転記変換における データ間・個人間のゆれに関する分析

土屋 智行 (国立国語研究所言語資源研究系)<sup>†</sup>

伝 康晴 (千葉大学文学部/国立国語研究所言語資源研究系)

小磯 花絵 (国立国語研究所理論・構造研究系)

## **Towards Automatic Transformation into CA Transcript Conventions: Differences of Transcript Strategy between Data and between Transcribers**

Tomoyuki Tsuchiya (Dept. Corpus Studies, NINJAL)

Yasuharu Den (Faculty of Letters, Chiba University/Dept. Corpus Studies, NINJAL)

Hanae Koiso (Dept. Linguistic Theory and Structure, NINJAL)

### 1. はじめに

近年、大規模な書き言葉コーパスの発展が見られる一方、話し言葉コーパスは音声収録・転記という初期段階に大きな負担がかかることが課題となって進展が遅れており、とくに大規模な会話コーパスは未着手である。国立国語研究所独創・発展型共同研究「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・2011年11月～2014年10月)は、既存の会話コーパスを共有化することで、この課題を解決することを目的として立ち上げられた。

既存のコーパスを共有するにあたって、転記方式の不統一や基本情報アノテーションの欠如といった問題がある。伝ほか(2012)は、本プロジェクトのメンバーが有する10数種のコーパスで用いられている転記方式を調査し、それらがCSJ方式と会話分析方式に概ね大別できることを示した。土屋ほか(2012)は、CSJ方式の韻律ラベルから会話分析方式の音調マーカへの自動変換を試み、会話データ/転記者ごとに言語特徴を重視するか、音響特徴を重視するかという転記方略の違いが存在し、この方略の違いに対処する必要があると論じた。ここでは、CSJ方式の言語・音響情報から会話分析方式の音調マーカを予測するための多変量モデルを構築することで、音調マーカの予測に貢献する言語・音響特徴を分析した。その結果、予測に貢献する言語・音響特徴が会話データ/転記者ごとに異なることが分かった。たとえば、あるデータ/転記者では、アクセント句の末尾単語や次末単語の品詞といった言語的な特徴が予測に強く貢献していたが、もう一方のデータ/転記者では、アクセント句末のF0値などの音響的な特徴のほうが貢献していた。しかし、この研究では、2人の転記者がそれぞれ1つずつのデータしか転記しておらず、上記の違いが会話データ自体の質的な違いに起因するのか、転記者の一般的な転記方略の違いに起因するのかがわからない。

本研究の目的は、この転記方略のゆれが、データ間・個人間でどのように現れるのかをより詳細に検討することである。そのため、音調マーカを予測するための言語・音響特徴の貢献

---

<sup>†</sup> ttsuchiya@ninjal.ac.jp

度の違いが、会話データの特徴によって生じるのか、あるいは転記者による転記方略の違いによって生じるのかを明らかにする。具体的には、土屋ほか (2012) で用いた 2 人の転記者による各 1 つずつの会話データに加え、これら 2 つの会話データをもう 1 人の転記者によって転記したデータも比較することで、データ間および個人間による転記方略の違いを分析する。

## 2. 方法

### 2.1 談話資料

本研究で用いる会話コーパスは、土屋ほか (2012) と同じ千葉大学 3 人会話コーパス (Den and Enomoto 2007) の 2 会話 (chiba0232 と chiba0432)、合計約 20 分である。本コーパスには、簡略版 CSJ 方式による転記テキストと、発話単位・形態論情報・韻律情報などの種々のアノテーションが与えられている。

### 2.2 転記・アノテーション

#### 2.2.1 CSJ 方式

CSJ 方式の転記テキストの例を図 1 に記す。この例には参考のために句末境界音調が記されているが、実際のアノテーションでは、これらは別ファイルとして用意され、時間情報などを利用し相互にリンクがとれる形で蓄積されている。

CSJ 方式では、X-JToBI (五十嵐ほか 2006) に基づく韻律情報が提供され、アクセント句の末尾に句末境界音調が付与される。本データでは、(1) 下降調 (L%) に加え、複合境界音調として、(2) 単純な上昇調 (L%H%)、(3) 上昇前に一定期間低ピッチが見られる上昇調 (L%LH%)、(4) 上昇下降調 (L%HL%) の 4 種類を区別した。下降調 L% は、複合境界音調が生じないアクセント句末に付与される音調であり、必ずしも明示的な下降が生じているわけではない。この点において、会話分析方式のピリオドとは若干異なる。また上昇調 L%H%、L%LH% も、疑問上昇調だけでなく強調上昇調なども含まれており、クエスチョンとは必ずしも一致しない。

#### 2.2.2 会話分析方式

会話分析方式の転記テキストの例を図 2 に示す。会話分析方式の転記に使われる種々の転記シンボルのうち、本研究では、ピリオド (per) ‘.’、クエスチョン (ques) ‘?’、コンマ (com) ‘,’、アンダーバー (ub) ‘\_’ の 4 つの音調マーカーに注目した。これらのマーカーはそれぞれ下降・上昇・継続・平坦の音調を表す。

会話分析方式による転記は、Gail Jefferson の体系 (Jefferson 2004) に準拠して、会話分析の研究者 3 名 (X 氏、Y 氏、Z 氏) によって行なわれた。X 氏は chiba0232 を、Y 氏は chiba0432 を転記し、Z 氏は chiba0232 と chiba0432 の両方を転記した。2013 年現在で、X 氏は約 7 年、Y 氏と Z 氏は約 6 年の会話分析経験を有する。以下に、X 氏、Y 氏、Z 氏それぞれの会話分析経験の概要を示す。

X 氏は、2003 年からカリフォルニア大学ロサンゼルス校で会話分析を学び、2006 年から本格的に会話分析による研究を行なっている。2010 年に日本に帰国した後も、各科研プロジェクトや研究会、データセッションへの参加を継続的に行なっている。また、会話分析以外に音声学 (イントネーション) の授業を受けた経験がある。

### CSJ 方式

281.7240 283.4033 B: 松下にじゃねえ (L%) 松下だろ (H%)  
283.4775 283.8650 B: 〈笑〉  
284.3166 285.3986 C: あれ (L%) もともと一緒なの (L%)  
285.5721 286.2149 B: もともと一緒 (L%)  
285.7004 286.1680 A: そうだよ (L%)

図 1 CSJ 方式の転記テキスト (括弧内の句末境界音調は別ファイル)

### 会話分析方式

B: 松下にじゃねえ, 松下だろ::.  
(0.5)  
C: あれ, もともと一緒なの?  
A: そう[だよ.  
B: [もともと一緒.

図 2 会話分析方式の転記テキスト

Y 氏は、2006 年から 2007 年にわたり語用論や談話分析の教科書を通じて会話分析の概念に触れ始め、2007 年からデータセッションへの参加や、独自に収録したデータおよび CSJ の転記を始めている。2008 年からは、カリフォルニア大学サンタバーバラ校で会話分析の授業を受け、2009 年から十数時間程度のデータの収録および転記を行なっている。会話分析以外にも、談話分析の専門的知識を有し、Du Bois 流の記法を学んでいる。

Z 氏は、2004 年から 2007 年までカリフォルニア大学ロサンゼルス校で会話分析を学び、2007 年以降は日本国内のデータセッションや研究会に参加している。大学院博士課程在籍時より、主要な分析手法の 1 つとして会話分析を採用している。また、会話分析以外に、認知言語学と談話機能主義言語学の知識を有している。

## 2.3 言語・音響特徴

CSJ の言語・音響情報から会話分析方式の音調マーカーを予測する多変量モデルでは、分析対象アクセント句から以下の言語・韻律特徴を抽出し用いた\*1。

### ■言語特徴

句末境界音調 (tone) アクセント句末の句末境界音調。L%・H%・HL%・LH%。

末尾単語の品詞 (lastPOS) アクセント句末尾の単語の品詞。品詞は以下の 7 種に分類した。体言・用言・助動詞・終助詞・接続助詞・その他の助詞・その他の品詞。

次末単語の品詞 (penultPOS) アクセント句の最後から 2 番目 (次末) の単語の品詞

### ■音響特徴

アクセント句の最小 F0 (f0MinAP) アクセント句中の F0 の最小値 (標準化得点)

\*1 音響特徴として他の特徴も抽出したが、これらの特徴との相関が高いため用いなかった。

アクセント句の最大 F0 (f0MaxAP) アクセント句中の F0 の最大値 (標準化得点)  
 句末単語の最大 F0 (f0MaxWord) 末尾単語中の F0 の最大値 (標準化得点)  
 アクセント句の最大パワー (pwrMaxAP) アクセント句中のパワーの最大値 (標準化得点)  
 句末単語の最大パワー (pwrMaxWord) 末尾単語中のパワーの最大値 (標準化得点)  
 アクセント句の平均モーラ長 (amdAP) アクセント句の継続時間をモーラ数で除したもの  
 (標準化得点)  
 最終抽出可能 F0 点の値 (lastF0Val) アクセント句中で最後に抽出できた F0 点の値 (標準化得点)

最終抽出可能 F0 点の位置 (lastF0Loc) 上記 F0 点の句末から計った時間 (対数値)

F0 と平均モーラ長は対数変換後、パワーはそのままで、話者ごとに標準化得点に変換した。

■その他の特徴 以上に加え、アクセント句自体の位置に関する以下の特徴を用いた。

発話冒頭からの位置 (loc) 発話単位中で先頭から何番目のアクセント句か (対数値)

発話末尾からの位置 (revLoc) 発話単位中で末尾から何番目のアクセント句か (対数値)

## 2.4 分析手順

データ間・個人間での転記方略のゆれを図 3 のように比較した。まず、データ内・個人間の比較として、chiba0232 では X 氏と Z 氏による転記の音調マーカ―を、chiba0432 では Y 氏と Z 氏による転記の音調マーカ―をアクセント句単位で整列し比較した (実線矢印)。音調マーカ―が付与されていないアクセント句は none のラベルを与えた。

次に、個人内・データ間の比較として、Z 氏によって転記された chiba0232 と chiba0432 の音調マーカ―を比較した (破線矢印)。この比較のために、CSJ 方式の転記から会話分析方式の転記へ変換する多変量モデルを一方のデータを学習データとして構築し、他方のデータの音調マーカ―の予測結果と人手による音調マーカ―とを比較した。多変量モデルとしてランダムフォレスト法 (Breiman 2001) を用い、統計解析ソフト R 言語の randomForest パッケージを使ってモデルを構築した (mtry = 4 とした)。アクセント句単位でデータを分節化し、2.3 で述べた言語・音響特徴を説明変数に用いた。

さらに、各データ・転記者の転記方略を検討するため、全 4 データに対してランダムフォレスト法による多変量モデルを構築し、OOB 推測に基づく各特徴の貢献度を推定した。

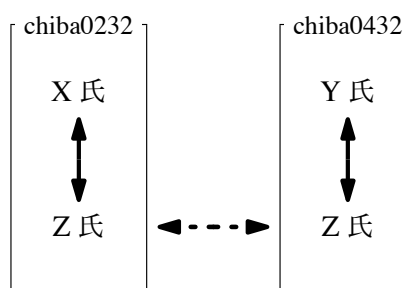


図 3 データ間・個人間比較

表1 データ内・個人間のゆれ

chiba0232 (一致率 = 76.0%、 $\kappa = .66$ )						chiba0432 (一致率 = 69.9%、 $\kappa = .58$ )					
Z氏						Z氏					
X氏	none	per	ques	com	ub	Y氏	none	per	ques	com	ub
none	<b>130</b>	12	4	2	1	none	<b>184</b>	8	0	0	0
per	9	<b>140</b>	15	0	0	per	30	<b>126</b>	1	3	0
ques	2	9	<b>58</b>	0	1	ques	5	11	<b>29</b>	1	0
com	33	20	2	<b>26</b>	0	com	92	14	1	<b>89</b>	5
ub	0	1	1	0	<b>1</b>	ub	0	13	0	0	<b>0</b>

表2 個人内・データ間のゆれ (転記者 = Z氏)

学習 = chiba0432、比較 = chiba0232 (正解率 = 72.8%、 $\kappa = .58$ )						学習 = chiba0232、比較 = chiba0432 (正解率 = 74.8%、 $\kappa = .59$ )					
観測値						観測値					
予測値	none	per	ques	com	ub	予測値	none	per	ques	com	ub
none	<b>134</b>	10	3	2	1	none	<b>272</b>	22	6	35	5
per	28	<b>164</b>	58	3	2	per	34	<b>129</b>	13	12	0
ques	0	3	<b>19</b>	0	0	ques	3	18	<b>12</b>	1	0
com	12	5	0	<b>23</b>	0	com	2	3	0	<b>45</b>	0
ub	0	0	0	0	<b>0</b>	ub	0	0	0	0	<b>0</b>

### 3. 結果

#### 3.1 データ内・個人間でのゆれ

chiba0232 における X 氏と Z 氏による音調マーカの対応、および chiba0432 における Y 氏と Z 氏による音調マーカの対応を表 1 に示す。全般的に、X 氏と Z 氏のほうが一致が高く (一致率 = 76.0%、 $\kappa = .66$ )、Y 氏と Z 氏では一致がより低かった (一致率 = 69.9%、 $\kappa = .58$ )。件数の少ないアンダーバー (ub) を別にすると、X 氏/Y 氏がコンマ (com) を付与している箇所 Z 氏がそうしていない例が多く見られた。また、chiba0432 では、Z 氏が音調マーカを付与していない箇所 (none) に Y 氏がピリオド (per) やコンマ (com) を付与している例も多く見られた。

#### 3.2 個人内・データ間でのゆれ

Z 氏による chiba0232 と chiba0432 の転記方略の違いを、多変量モデルによる予測結果と人手ラベルとの対応によって調べた。chiba0432 を学習データとして構築した多変量モデルによる chiba0232 の予測結果と Z 氏による人手ラベルとの対応、および chiba0232 を学習データとして構築した多変量モデルによる chiba0432 の予測結果と Z 氏による人手ラベルとの対応を表 2 に示す。正解率 (一致率) はいずれも比較的高いが (chiba0232 : 72.8%、chiba0432 : 74.8%)、いずれの方向の予測でも、クエスチョン (ques) をピリオド (per) と誤って予測する例が多く、chiba0432 ではコンマ (com) をラベルなし (none) と誤って予測する例も多かった。

表3 個人内・データ間のゆれ (転記者 = Z 氏) (句末境界音調ごと)

学習 = chiba0432、テスト = chiba0232 L% (正解率 = 82.3%、 $\kappa = .67$ )						学習 = chiba0232、テスト = chiba0432 L% (正解率 = 80.5%、 $\kappa = .59$ )					
予測値	観測値					予測値	観測値				
	none	per	ques	com	ub		none	per	ques	com	ub
none	<b>118</b>	8	1	0	1	none	<b>254</b>	20	1	20	5
per	17	<b>109</b>	15	1	1	per	25	<b>96</b>	1	5	0
ques	0	0	<b>0</b>	0	0	ques	1	7	<b>0</b>	0	0
com	5	0	0	<b>0</b>	0	com	0	0	0	<b>0</b>	0
ub	0	0	0	0	<b>0</b>	ub	0	0	0	0	<b>0</b>

学習 = chiba0432、テスト = chiba0232 H% (正解率 = 57.8%、 $\kappa = .36$ )						学習 = chiba0232、テスト = chiba0432 H% (正解率 = 54.2%、 $\kappa = .32$ )					
予測値	観測値					予測値	観測値				
	none	per	ques	com	ub		none	per	ques	com	ub
none	<b>15</b>	2	1	2	0	none	<b>15</b>	0	5	5	0
per	6	<b>41</b>	33	0	0	per	6	<b>21</b>	9	0	0
ques	0	1	<b>11</b>	0	0	ques	2	11	<b>9</b>	0	0
com	4	0	0	<b>0</b>	0	com	0	0	0	<b>0</b>	0
ub	0	0	0	0	<b>0</b>	ub	0	0	0	0	<b>0</b>

この傾向をさらに詳しく見るため、予測結果と人手ラベルとの対応をアクセント句の句末境界音調ごとに細分化した (表3; 件数の少ないHL%とLH%は除く)。いずれの方向の予測でも、L%と比べてH%での正解率がかなり低かった (chiba0232: 82.3% vs. 57.8%、chiba0432: 80.5% vs. 54.2%)。とくに、H%の典型的な機能であるクエスチョン (ques) を正しく予測できていない例が多かった。

### 3.3 言語・音響特徴の貢献度

音調マーカの予測に貢献する言語・音響特徴を図4に示す。

まず、データ内・個人間で比較すると (列方向)、chiba0232のX氏とZ氏の比較 (左列) では、上位6位までの特徴 (revLoc, tone, lastF0Loc, lastPOS, lastF0Val, f0MaxWord) の相対的な貢献度が同じであった。一方、chiba0432のY氏とZ氏の比較 (右列) では、上位7位までに含まれる特徴 (revLoc, tone, lastF0Loc, lastF0Val, f0MinAP, amdAP, lastPOS) は同じであるが、その順序は異なった。Y氏はamdAPをより重視する傾向があり、Z氏はlastPOSをより重視する傾向があった。

次に、個人内・データ間で比較すると (下段行方向)、Z氏のchiba0232とchiba0432の比較において、上位7位までに含まれる特徴のうち6個 (revLoc, tone, f0MinAP, lastF0Loc, lastF0Val, lastPOS) は共通していたが、その順序は異なった。chiba0232ではlastF0LocやlastF0Valをより重視する傾向があり、chiba0432ではlastPOSをより重視する傾向があった。また、両者に共通しない特徴として、chiba0232ではf0MaxWordが、chiba0432ではamdAPが重視されていた。

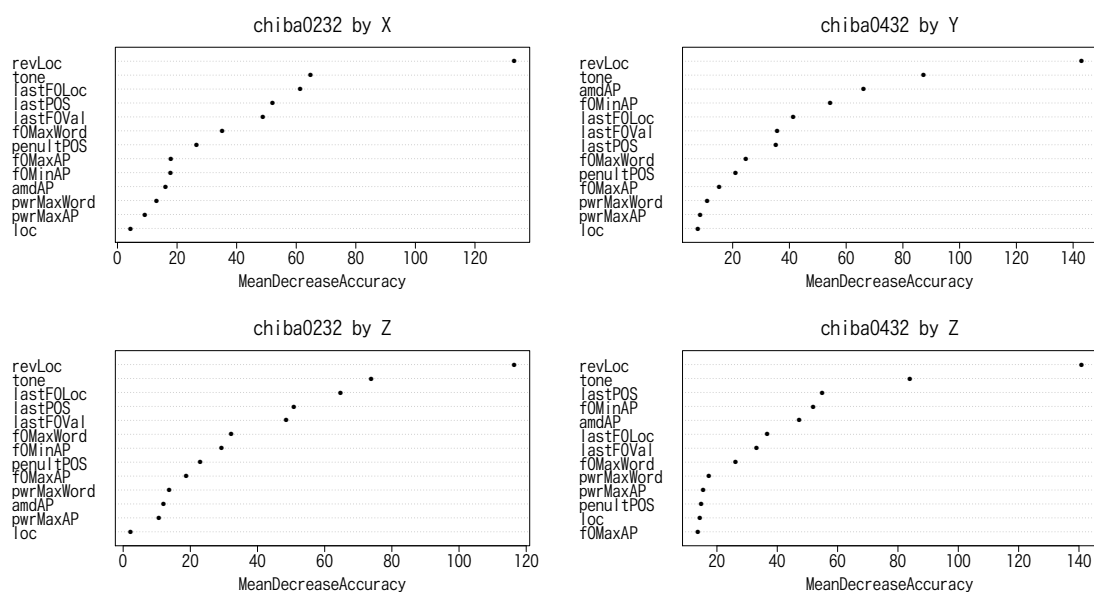


図4 言語・音響特徴の重要度

#### 4. 議論

転記者の音調マーキングの全般的な傾向として、Z氏がX氏、Y氏よりも音調マーカを付与していない箇所が多い。chiba0232のアクセント句全体におけるnoneの比率は、X氏が31.9%であったのに対して、Z氏は37.3%と少し高かった。同様にchiba0432では、Y氏の31.4%に対して、Z氏は50.8%に対して音調マーカを付与しなかった。とくにY氏とZ氏の違いが大きいが、この点はY氏の音調マーキングの手順に起因するものと思われる。Y氏はカリフォルニア大学サンタバーバラ校でDu Bois流の転記記法を学んだ経験があり、その影響から、まず最初に転記テキスト中にイントネーションユニット (Du Bois et al. 1993) を同定し、イントネーションユニット末ごとに音調マーカを付与するという方略を採用していた (土屋ほか2012)。このため、Z氏よりも音調マーカを多く付与する傾向があったものと思われる。

また、転記者間で異なる音調マーカを付与している例も散見された。大きな違いとして、Z氏がX氏、Y氏よりもコンマ (com) を付与する箇所が少ないという点が挙げられる。chiba0232でコンマ (com) を付与した箇所はX氏の81箇所に対してZ氏は28箇所と少なく、同様にchiba0432ではY氏の201箇所に対してZ氏は93箇所と少ない。このように、どの箇所を継続音調とするかについては、転記者ごとの違いが強く現れると考えられる。

以上のデータ内・個人間のゆれに関しては、音調マーカを付与する際に重視する言語・音響特徴の違いという点に一部原因を求めることができる。X氏やZ氏はlastPOSのような言語特徴をより重視していたのに対して、Y氏はamdAPのような音響特徴をより重視していた。このことは上に述べた音調マーキングの手順の違いの現れと思われ、結果として個人間のゆれの一因となっているようである。ただし、X氏とZ氏は重視する特徴が似通っており、両者の継続音調のマーキングの違いは他の何らかの要因に起因するものと思われる。

さらに、個人内・データ間の違いとして2つの会話データ間でのZ氏の音調マーキングを比較すると、音調マーカの予測に貢献する言語・音響特徴は全般的には共通するものの、細部では一部異なっていた。とくに chiba0432 では amdAP がより重視されており、この傾向は同じデータに対するY氏の転記方略と共通していた。このことから、音調マーキングを行なう際に重視する特徴はデータごとに異なると考えられる。これには、会話データの話者の言語的な特徴が関係していると思われる。chiba0432 では、話者の一人は関西方言を話しており、継続や下降の音調が他の話者と異なっている。また、事後インタビューによると、Z氏は軽微な上昇の後に大きな下降が起きる箇所にも継続音調のマーカを付与していると答えているが、chiba0432 の話者の一人は軽微な上昇の直後にとくに大きな下降を用いることが多いという点などから、発話の終了や継続を示す話者の発話方略にも影響を受けていると考えられる。

最後に、句末境界音調ごとの多変量モデルの予測精度は、L% と比べて H% でかなり低かった。とくに、H% の典型的な機能であるクエスチョンを正しく予測できていない例が多かった。CSJ 方式で採用している X-JToBI の上昇調 L%H% には疑問上昇調だけでなく強調上昇調なども含まれており、ここから会話分析方式のクエスチョンを抽出するには現状の言語・音響特徴だけでは難しいことがわかった。今後、韻律ラベリング方法も含め再検討する必要がある。

以上のように、CSJ 方式から会話分析方式への転記変換において、データ間・個人間でさまざまなゆれが存在することがわかった。とくに、このゆれは句末境界音調や音調マーカごとに大きく異なっていた。今後は、転記方略の違いをモデル化し、予測精度を上げる必要がある。

謝辞 会話分析方式の転記を作成していただいた遠藤智子・黒嶋智美・横森大輔の各氏に感謝します。本研究は国立国語研究所独創・発展型共同研究「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)による成果である。

#### 参考文献

- Breiman, Leo (2001). "Random forests." *Machine Learning*, 45, pp. 5–32.
- Den, Yasuharu, and Mika Enomoto (2007). "A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation." Toyoaki Nishida (Ed.), *Conversational informatics: An engineering approach*. Hoboken, NJ: John Wiley & Sons. pp. 307–330.
- 伝康晴・土屋智行・小磯花絵 (2012). 「多様な様式を網羅した会話コーパスの共有化」 第1回コーパス日本語学ワークショップ予稿集, pp. 227–234.
- Du Bois, John W., Stephan Schuetze-Coburn, Susanna Cumming, and Danae Paolino (1993). "Outline of discourse transcription." Jane A. Edwards, and Martin D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research*. Hillsdale, NJ: Lawrence Erlbaum. pp. 45–89.
- 五十嵐陽介・菊池英明・前川喜久雄 (2006). 「韻律情報」 『国立国語研究所報告 124: 日本語話し言葉コーパスの構築法』 pp. 347–453.
- Jefferson, Gail (2004). "Glossary of transcript symbols with an introduction." Gene Lerner (Ed.), *Conversation analysis: Studies from the first generation*. Amsterdam/Philadelphia: John Benjamins. pp. 13–31.
- 土屋智行・伝康晴・小磯花絵 (2012). 「会話コーパスの転記方式の相互変換に向けて—イントネーションに着目して—」 第2回コーパス日本語学ワークショップ予稿集, pp. 117–126.

#### 関連 URL

「会話コーパス」ホームページ：<http://www.jdri.org/kaiwa/>