

非母語話者と母語話者のディベート会話における母語話者 テキスト入力の影響

宋暁宇^{†1} 埴裕美^{†1} 井上智雄^{†2}

概要: 第二言語会話における母語話者 (NS) と非母語話者 (NNS) の会話では意思疎通が困難な場合が見られる。このような場面を支援するために、NS がキーワードなどをテキスト入力し、それをリアルタイムで NNS と共有する手法がこれまでに提案され、参加者間の共通理解が増えるなど一定の有用性が確認されている。本稿では、この手法をディベート会話に適用した場合の NS と NNS の発話行動について、NS がテキストを入力している区間とそうでない区間とを分けて比較した。その結果、NS はテキスト入力により自身の発話行動が変わらず、NS にとって会話しながらテキスト入力する時の負荷やストレスはさほど大きくないことがわかった。一方、NNS は NS テキスト入力時の発話時間が少なく、発話長が短いことがわかった。

キーワード: 第二言語、会話支援、音声会話、テキスト入力、視線

Effects of native speaker's text input in the debate-style conversation with a non-native speaker

XIAOYU SONG^{†1} HIROMI HANAWA^{†1} TOMOO INOUE^{†2}

1. はじめに

近年、世界中でグローバル化が進んでおり、母語の異なる人々の間でコミュニケーションを行う機会が増加している[1]。しかし、母語ではない言語を話す非母語話者(以下 NNS)にとって、母語話者(以下 NS)と音声会話を行うのは、NS と NNS の言語の流暢さに隔たりがあるため必ずしも容易ではない[2]。

言語の不均衡を解消する方法の一つとして、文字情報提示による支援が研究され、これまでにテキスト提示の方法によってコミュニケーションや理解に対して異なる効果や影響があることが示されている[3-6]。機械翻訳による翻訳文は誤りが多いため、NS が翻訳文中の重要部分を強調表示する方法[3]が提案され、聴覚障害者の講義参加のための字幕は早く正確な提示が求められることが知られている[4]。

これらの先行研究を踏まえ、これまでに会話中の NNS の負担を考慮して、NS が NNS との会話中に、NNS にとって重要な部分や理解しにくい部分を NS がキーボードでテキスト入力し、それを NS と NNS に表示するテキスト入力会話支援方法(以下、NS テキスト入力会話)が検討されている[7][8][9]。その結果参加者の相互理解と会話後の一致した記憶が増加する効果が示されている[7][8]。

この、NS テキスト入力会話において NS がテキストを入

力している区間(以下、入力区間)とそうでない区間(以下、非入力区間)がある。本研究では、入力区間と非入力区間における差異を分析した。

2. 関連研究

2.1 NNS を含む音声会話の支援

これまでの研究では、音声会話は音声品質と個人の言語能力に依存しているため、対面会話より難しいとされている[10]。特に、母語ではない言語を話す NNS にとって、NS と音声会話を行うのは、必ずしも容易ではない[11]。そこで、NNS を含む音声会話の支援が研究されてきた。

音声会話における NNS の理解や会話をサポートするため、自動音声認識(ASR)や機械翻訳(MT)等の技術を使う研究がされた。Pan らは英語の映像(音声)に ASR により字幕を付与し、視聴した NNS の理解度を調査した。結果として、10%以内の誤字率と 2 秒以内の遅延であれば NNS の理解度に役に立つ。一方、ASR による誤字を含む字幕は、完璧な字幕と比べて NNS の理解を妨げることが示された[5]。

NNS の負担を考慮し、Gao らは多言語会話において MT を用いて重要なテキストをハイライトにして提示することで、MT エラーを避け、重要部分が明瞭になり協調作業が向上したとしている[4]。また、山下らは NNS の負担を低減するため、ASR の出力文を見て NS が会話中の重要部分をハイライ

^{†1} 筑波大学大学院図書館情報メディア研究科
Graduate School of Library, Information and Media Studies, University of Tsukuba.

^{†2} 筑波大学大学院図書館情報メディア系
Faculty of Library, Information and Media Science, University of Tsukuba

トに表示する工夫を行なった。NNS がより明快、快適と感じるとの結果が得られている[12]。これらの研究で、音声会話のテキスト会話全文ではなく、重要な部分を提示することにより NNS の会話を支援できることがと示された。

2.2 NS によるテキスト入力手法の効果

話しながらテキストを書くのは新しいことではない。病院において患者に理解しやすく説明するため、医療用カウンセリングでは医師が紙に書きながら話す方法が行われていた[13]。そして、Chapanis は 70 年代に音声+テキスト入力の研究を始めている[6]。

山下らはテキスト表示の効果を最大にするため、NNS と会話中に、NS は ASR の出力文を見ながら NS 会話中の重要部分をハイライトに表示する手法を提案した[12]。また、井上らは NNS を含む音声会話において、NNS の会話支援と NS 作業負担を軽減するための手法として、NS が話しながら会話中のキーワードをタイピングし、それを NNS と共有する方法を提案した[9]。

提案手法を用いた会話では、会話中に共通基盤を示す語句が増えること、会話後に確認した内容が一致した語句の増加が確認された[7][8]。また、参加者の理解しやすさや文字の見やすさなどで高く評価され、インタビューでも理解しやすくなり不安や緊張が軽減されたという裏付けを得た[7][8]。しかし、これまでの研究は、NS テキスト入力作業は話者の発話行動への影響が調べられていない。

2.3 ながら会話

日常的に、人々が会話をする時、会話に専念している時ばかりではない。何らかの他の作業と並行して行う会話、いわゆる「ながら会話」をしている場合が多い[14]。並行作業やマルチタスクでは作業中の会話が多いため、井上らは、食事がある会話とない会話における多人数会話の発話とジャスチャを計量的に分析した。結果として食事を伴う会話は参加者間の会話行動差異を減少し、平準化する効果があることがわかった[14]。また、オンラインゲーム中の会話を対象とした研究では、ゲーム内の行動そのものが発話に取って代わることが分かった[15]。別な例では、箕輪らは運転中の助手席から運転者への情報提供という目的で、運転手が助手席者と会話する実験を行なった。結果として運転中の会話は道路情報により発話行動が変わるということが分かった[16]。以上から、ながら作業によって会話への影響が異なることが分かる。しかし、NS テキスト入力という作業が会話へどのように影響するかを調べた研究が見当たらない。そこで本研究では入力区間と非入力区間を分けて NS と NNS それぞれの発話を分析することで、NS テキスト入力という会話中のながら作業が NS と NNS の発話行動に与える影響を調べた。

3. 方法

3.1 データ収集

埴ら[7][8]は NS テキスト入力会話の評価実験を行った。評価実験の入力条件では NS がテキスト入力を行い、NNS に対してテキストを提示しながら NS と会話した。対照条件では、NS が入力を行わず、NNS と会話した。原発問題と死刑問題に関するディベート形式の会話を図 1 の環境で行った。各条件には 7 分間の会話を行った。参加者は初対面の NS16 名と NS16 名による 16 ペアであった。NNS はすべて日本語能力検定試験 1 級合格者であった。

本研究では、評価実験の入力がある会話（入力条件）のデータを分析対象とした。



図 1 実験環境

図 2 は参加者使用モニタである。モニタ画面の左側部分に NS はキーボードでテキストを入力し。モニタ右側にディベート会話用の賛成反対の参考資料が配置され、モニタ右側は会話相手に共有されなかった。

3.2 分析項目



図 2 参加者使用 PC モニタの画面

NS テキスト入力会話において入力区間と非入力区間とを分けて、NS と NNS の発話量と NNS の視線を調べた。入力区間と非入力区間の時間割合、また発話量について、発話時間、発話頻度、発話長を分析した。発話時間とは話者ごとの 1 分あたりの発話時間を示す。発話頻度とは 1 分あたりにテキスト入力区間に発話開始する回数である。発話

長とは一回の発話をする時間の長さである。視線については入力区間と非入力区間の NNS の視線を、入力されたテキスト、参考資料、よそ見等について分類した。

3.3 データ処理

収集したデータは NS と NNS それぞれを撮影した横側ビデオ、全体を撮影したビデオ、モニタ画面のキャプチャビデオである。本研究ではペア 16 組の 7 分間（会話開始からの 7 分間）の会話のデータ（合計 112 分）を分析した。各会話について NS と NNS それぞれを横から撮影したビデオ 2 つとモニタ画面のキャプチャビデオ一つを時間的に同期させ、ビデオ（図 3-A）を作成した。作成したビデオはビデオ分析ツール ELAN[17]を用いて NS の発話と NNS の発話、NS の入力及び NNS 視線をラベリングした。

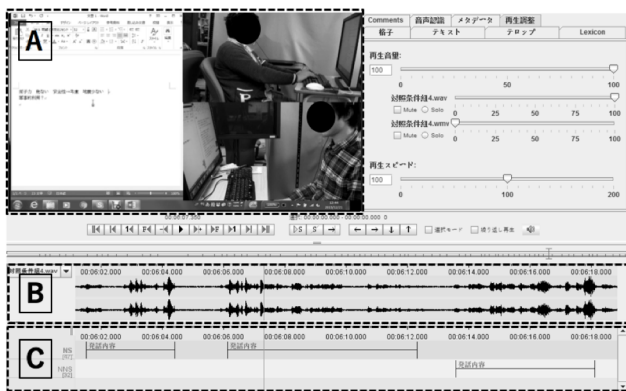


図 3 ELAN を用いたラベリングの様子：会話中の映像データ (A)，音声会話の波形(B)，ラベリングした結果(C)

発話の有無は、まず 300ms 以上の無音区間によって分割される単一話者の連続する音声区間を発話とし（間休止単位：Inter Pausal Unit (IPU)），ELAN のセグメンテーション機能を用いて発話区間を自動分割した[18]。次に自動分割された発話区間が機能しない箇所1を実験者が手作業で修正した。

以下は視線のラベリングの手順を示す。視線変化を把握するため分析者のビデオ観察により参加者の視線行動を以下の種類に分類した。



図 4 NNS 視線変化

1. 入力されたテキストを見る（モニタ左側：図 4-a）
2. 参考資料を見る（モニタ右側：図 4-b）
3. よそ見する（モニタ以外の横，下，上等見る：図 4-c）

1雑音が入っている箇所(全体の約 5%)

4. その他（判断できない時）

以上の基準により実験者二人が独立にカウントした。

図 3 はラベリングの様子を示す。A の部分は会話中の映像データを示す。B の部分は音声会話の波形と C の部分はラベリング結果を示す。

NS の発話と NNS の発話，NS の入力及び NNS の視線のラベリング例を図 5 に示す。

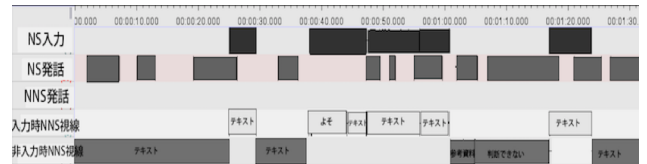


図 5 NS の発話と NNS の発話，NS の入力及び NNS の視線のラベリング例

上から 1 段目は NS 入力内容と時間，上から 2 段目は NS 発話内容と発話時間，上から 3 段目は NNS 発話内容と発話時間，上から 4，5 段目は NNS の視線が記録される。

NS 入力区間に発話開始が含まれる発話は入力区間の発話とし，それ以外の発話は非入力区間の発話とした。

4. 結果

4.1 入力区間と非入力区間の量と比率

表 1 に入力区間と非入力区間の 1 分あたりの量およびその比率を示す。入力がある会話では全体の 22.1%が NS 入力する時間であり，全体の 77.9%が入力していない時間であった。この差について Wilcoxon の符号付順位検定の結果，有意差が認められた ($Z=-3.521, p<0.01$)。

表 1 入力区間および非入力区間の量とその比率

	入力時	非入力時
1分あたり 時間 (秒)	13.26	46.74
比率 (%)	22.1	77.9

4.2 入力区間と非入力区間の発話量

入力区間と非入力区間の発話時間，発話頻度，発話長について，NS と NNS とを分けた。結果を表 2 に示す。

表 2 NS・NNS 毎の発話量

	入力時		非入力時	
	NS	NNS	NS	NNS
発話時間 (秒)	21.39	13.31	23.25	19.78
発話頻度 (回)	9.81	7.91	9.63	7.61
発話長 (秒)	2.18	1.68	2.41	2.60

4.2.1. 発話時間

図 6 に入力区間と非入力区間 1 分当たりの発話時間を示す。NS の場合、Wilcoxon の符号付順位検定の結果、入力区間と非入力区間において発話時間の差があるとはいえなかった ($Z=-1.589$, $p=0.113$)。NNS の場合は、非入力区間は入力区間よりも発話時間が有意に多かった ($Z=-4.516$, $p<0.01$)。従って、NS がテキスト入力する時、NNS の 1 分当たりの発話時間が有意に少なかった。

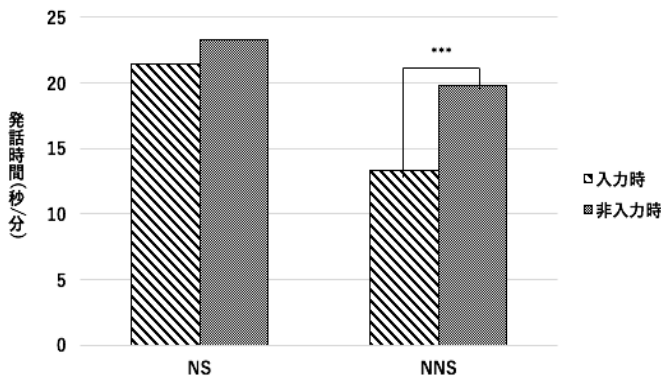


図 6 発話時間

4.2.2. 発話頻度

図 7 に発話頻度を示す。NS について、Wilcoxon の符号付順位検定の結果、入力区間と非入力区間の発話頻度に差があるとはいえなかった ($Z=-1.273$, $p=0.102$)。NNS についても同様であった ($Z=-1.422$, $p=0.154$)。

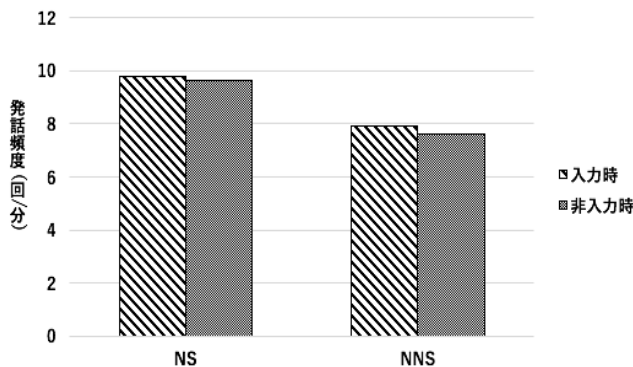


図 7 発話頻度

4.2.3. 発話長

図 8 に発話長を示す。NS の場合、Wilcoxon の符号付順位検定の結果、入力区間と非入力区間において発話長の差があるとはいえなかった ($Z=-2.20$, $p=0.826$)。NNS の場合、入力区間の発話長は 1.68 秒、非入力区間の発話長は 2.60 秒であり、有意差が認められた ($Z=-6.954$, $p<0.01$)。

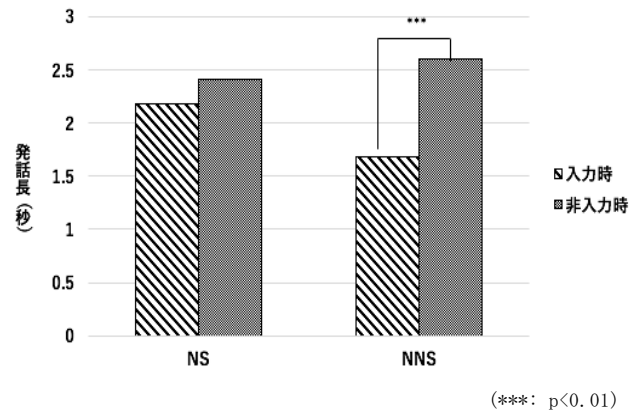


図 8 発話長

4.3 NNS の視線

図 9 と図 10 に入力区間と非入力区間における NNS の視線分布を示す (実験者 2 名カウントした結果の平均値)。実験者が独立にカウントした結果の一致度は十分高かった ($k=0.79$)。入力区間では NNS は入力されたテキスト (モニタ画面の左側部分) を見る時間が多かった。一方、非入力区間では NNS は特に決まった箇所を見ているわけではないということが分かった。

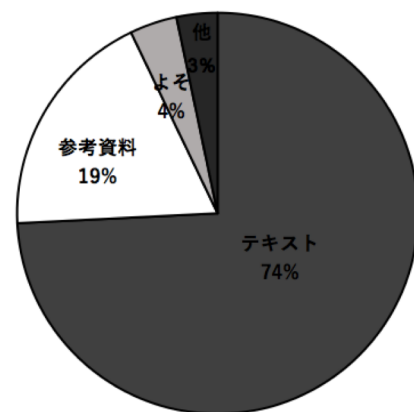


図 9 入力区間における NNS の視線分布

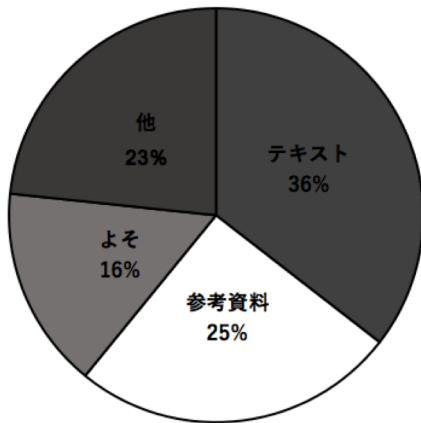


図 10 非入力区間における NNS の視線分布

4.4 入力区間の話者発話行動と NNS 視線の関係

表 1 に示したように、1 分あたりの入力区間は 13.26 秒であり、非入力区間は 46.74 秒である。入力区間について、さらに NNS が発話する場合と NS が発話する場合に分け、NNS の視線分布を調べた。表 3 のように、入力区間で NNS が発話する場合、NNS がテキストを見る時間は平均 1.81 秒であり、全体の 45% を占め、参考資料を見る時間は平均 1.99 秒であり、全体の 49% を占めている (図 11 左)。一方、入力区間で NS が発話する場合、NNS が入力されたテキストを見る時間は平均 4.74 秒であり、全部の 88% を占めた (図 11 右)。

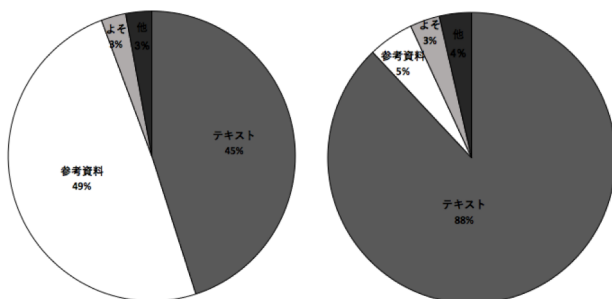


図 11 入力区間における NNS の視線分布 (左図: NNS 発話時
 右図: NS 発話時)

5. 考察

5.1 NS テキスト入力の NS 発話行動への影響

入力区間と非入力区間の割合は、入力区間 : 非入力区間 = 約 2 : 8 であった。図 6-8 に示したように、NS において入力区間と非入力区間の発話時間、発話頻度と発話長には差があるとはいえなかった。我々は何か作業をしながら会話する時に、しばしば作業により会話行動が制約される。例えば、食事によって発話やジェスチャが制約される [14]。しかし、NS は入力区間と非入力区間における発話行動の差

表 3 入力時 NNS, NS 発話する場合に NNS の視線 (1 分あたり)

	入力区間で NNS 発話	入力区間で NS 発話
テキストを見る時間 (秒)	1.81	4.74
参考資料を見る時間 (秒)	1.99	0.27
よそを見る時間 (秒)	0.11	0.19
その他	0.12	0.2

異が見られなかった。Eggemeier によれば、負荷が高いほど作業パフォーマンスが悪くなるという反比例関係が存在しているが、負荷レベルが増えても作業者の情報処理能力の範囲内であれば増加した作業負荷を補償する能力を有するため正解率や会話行動などの作業パフォーマンスが変わらない [19]。従って NS の発話行動に差が認められないということから、会話しながらテキスト入力する時の負荷やストレスはさほど大きくないと考えられる。

5.2 NS テキスト入力の NNS 発話行動への影響

図 6-8 に示したように、NNS は、入力区間において発話時間が有意に減り、発話長が有意に短くなった。言い換えれば NS テキスト入力は NNS の発話行動を制御した。会話中の NNS の視線について、入力区間に NNS は入力されたテキスト (モニタ画面の左側部分) を見る時間が多かった。NS が自分の発話に関連するキーワードを入力する時、NNS はほとんどキーワードを見ていた。また、箕輪ら [16] は運転中に助手席者はどのタイミングにどのような表現手法で運転手に道路情報を提供するかを調査する実験を行い、助手席者は車線変更をする車の量が多くタイミングが悪い時には運転者の運転負荷を考慮して余裕がないときは必要以上の情報提供を控え運転に集中させ、後で説明する場合が多いということを示している。本研究で入力区間において NNS は運転中の助手席者の立場になると考えられる。NNS が NS の作業負荷を考慮し、タイピングに集中させるため発話を控えたのではないかと考えられる。

6. 限界

本研究ではディベートを会話テーマとして NNS と NS の音声会話における NS によるテキスト入力の影響を調べた。

しかし、ディベート会話による結果は自由会話とは異なる可能性がある。またコンピュータを介したコミュニケーションでは会話をマネージメントしたり、話し方や振る舞いの行動変化がよく見られる [20]。これらの点においてディベート会話が参加者の発話行動に影響を与えた可能性が

ある。また、提案手法の有効性が他の会話に適用するかどうかは疑問である。そのため、今後自由会話における提案手法の有効性を検討予定である。

7. まとめ

NS と NNS の会話で意思疎通が困難な場面を対象とした、非母語による会話支援のために、NS がテキスト入力し、それをリアルタイムで NNS と共有する手法が提案されている。本稿では、テキストの入力区間と非入力区間を分けて発話行動を調査した。結果として、NS には入力区間と非入力区間の発話行動の差異が見られなかった。一方、NNS は入力区間において発話時間が有意に減り、発話長が有意に短くなった。

文 献

- [1] 015 年訪日外客数総数出典：日本政府観光局 (JNTO)http://www.jnto.go.jp/jpn/reference/tourism_data/visitor_trends/
- [2] Novinger, T. Intercultural Communication: A Practical Guide, University of Texas Press, Austin, TX, USA, 2001.
- [3] Ge Gao, Naomi Yamashita, Ari MJ Hautasaari, Andy Echenique, and Susan R. Fussell. Effects of public vs. private automated transcripts on multiparty communication between native and non-native English speakers. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14), pp843-852. ACM, New York, NY, USA, 2014
- [4] Ge Gao, H-C. Wang, D., and S. R. Fussell. "Same translation but different experience: the effects of highlighting on machine-translated conversations." In Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI '13), pp. 449-458 Paris, Apr. -May 27, 2013
- [5] Pan, Y., Jiang, D., Yao, L., Picheny, M., & Qin, Y. Effects of automated transcription quality on non-native speakers' comprehension in real-time computer-mediated communication. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems pp. 1725-1734. ACM. New York, NY, USA, 2010, April
- [6] Chapanis, A. : Human Factors in Teleconferencing Systems. Final Report, John Hopkins University, Baltimore, Maryland. Department of Psychology, p. 53, 30 November 1976
- [7] Hiromi Hanawa, Xiaoyu Song, Tomoo Inoue, Key-Typing on Teleconference: Collaborative Effort on Cross-Cultural Discussion, Collaboration Technologies and Social Computing Volume 647 of the series Communications in Computer and Information Science, Proceedings of the 8th International Conference, (CollabTech 2016), CCIS 647, pp. 74-88, Kanazawa, Japan, September 14-16, 2016.
- [8] 塙裕美, 宋曉宇, 井上智雄, NS の文字入力による NNS との会話支援-NS による会話中のテキスト入力が音声会話に与える影響-, 電子情報通信学会研究報告, Vol. 116, No. 31, pp. 139-144, ヒューマンコミュニケーション基礎 HCS2016.
- [9] Tomoo Inoue, Hiromi Hanawa, Xiaoyu Song, With a little help from my native friends: A method to boost non-native's language use in collaborative work, Proceedings of the Ninth International Workshop on Informatics, pp 223-226, 2015.
- [10] Yamashita, N., Echenique, A., Ishida, T., and Hautasaari, A. "Lost in transmittance: how transmission lag enhances and deteriorates multilingual collaboration" Proc. of the conference on computer supported cooperative work, pp. 923-934, February 2013.
- [11] 山下直美, エチエニーケアンディ, 葛岡英明, 石田亨, ハウタサーリアリ 国際電話会議の負担を軽減する手法の提案と評価 情報処理学会論文誌 Vol. 54 No. 6 pp1794-1806 June 2013
- [12] Mei-Hua Pan, Naomi Yamashita, HaoChuan Wang, Task Rebalancing: Improving Multilingual Communication with Native Speakers Generated Highlights on Automated Transcripts Proceedings of the Conference on Computer Supported Cooperative Work and Social Computing(CSCW'17), pp310-321 ACM Portland, OR, USA February 25-March 1, 2017
- [13] 西阪仰, 高木智世, 川島理恵 女性医療の会話分析 ソキウス研究叢書: 6 文化書房博文社 2008. 6, ISBN 4830111283
- [14] 井上智雄 大武美香 多数人会話における食事の有無の影響-会話行動の平準化 ヒューマンインタフェース学会論文誌 Vol. 13, No. 3, 2011. p 195
- [15] McEwan, Gregor; and Carl Gutwin. Chess as a conversation: Artefact-based communication in online competitive board games. In GROUP'16. Proceedings of the 19th International Conference on Supporting Group Work, Sanibel Island, USA, pp. 21-30. New York, ACM, 13-16 November 2016.
- [16] 簗輪 要佑, 稲垣 和芳, 梶川 忠彦 (U'eyes Design Inc.) 北島 宗雄, 赤松 幹之 (独立行政法人 産業技術総合研究所) 北崎 智之, 黒田 浩一, 丸山 泰永 (日産自動車株式会社) ドライバーにとって気の利いた情報とは ~実走行時の運転者と同乗者の自然対話の調査分析~Symposium on Mobile Interactions 2008, 2008/7/3-4
- [17] ELAN<https://tla.mpi.nl/tools/tla-tools/elan/>
- [18] 榎本美香, 石崎雅人, 小磯花絵, 伝康晴, 水上悦雄, 矢野博之. 相互行ため分析のための単位に関する検討. 電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎, 104(445), pp45-50, 2004.
- [19] Eggemeier, F. T., Properties of work loadassessment techniques, in Hancock, P. A. and Meshkati, N. (Eds), HumanMentalWorkloadpp. 41-62Amsterdam:North-Holland 1988
- [20] Marie-Noelle Lamy Oral conversations online: Redefining oral competence in synchronous environments ReCALL Volume 16, Issue 2 November 2004, pp. 520-538