

同時通訳データベース

プロの同時通訳者の音声を収録した世界随一のデータベースです。音声・言語処理技術の開発の他、言語学、心理学、認知科学、通訳学、語学教育など、言語に関する多様な研究目的に利用することができます。

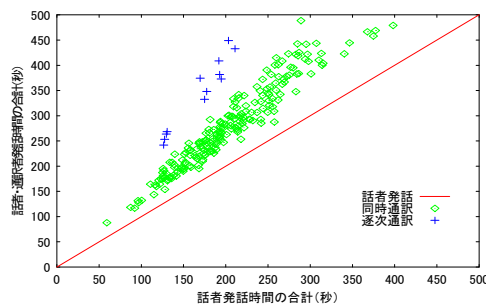
■ データベースの特徴 ■

<http://sidb.jp/>

- **独話と対話を大量に収録** 通訳対象として、日常的トピックを題材とした講演データ(独話)(通訳者(のべ)約220名)と旅行ドメインの会話データ(対話)(通訳者(のべ)約630名)を収集しています。いずれも10GBを超える大規模データとなっています。
- **英日・日英通訳音声を収集** 独話データベースには英語講演の英日通訳、日本語講演の日英通訳を、また、対話データベースには、英語話者と日本語話者による異言語間会話を収録しています。
- **同一講演に複数の通訳者** 一つの講演に対して、経験年数の異なる複数の通訳者を設定しています。通訳者による振る舞いの比較分析に利用できます。
- **発話時間情報を付与** 発話をポーズで分割し、開始時刻・終了時刻を付与しています。発話に対応付けることにより、訳出タイミングを取り出せます。



訳出タイミングの分析



大量データに基づく
統計的な解析が可能



通訳音声の収集状況

■ 価格 ■

独話データベース (DVD 3枚)	150,000 円 (アカデミック価格 30,000 円)
対話データベース (DVD 3枚)	150,000 円 (アカデミック価格 30,000 円)

*アカデミック価格は、研究者個人または大学の研究室での購入が対象です。



(公財) 名古屋産業科学研究所中部TLO
〒464-8603 名古屋市千種区不老町名古屋大学VBL棟4階
Tel:052-788-6010 Fax:052-788-6012 Email:ctlo@nisri.jp

同時通訳データベースに関連する 主な研究成果

- 松原 茂樹: 同時通訳の工学と科学 一次世代自動通訳技術の実現に向けて, 情報処理, Vol. 49, No. 6, pp. 617-623 (2008).
- Shigeki Matsubara, Akira Takagi, Nobuo Kawaguchi, and Yasuyoshi Inagaki: Bilingual Spoken Language Corpus for Simultaneous Machine Interpretation Research, Proceedings of 3rd International Language Resources and Evaluation Conference (LREC-2002), Vol. I, pp. 153-159 (2002).
- 松原 茂樹, 相澤 靖之, 河口 信夫, 外山 勝彦, 稲垣 康善: 同時通訳コーパスの設計と構築, 通訳研究(日本通訳学会誌), No.1, pp. 85-102 (2001).
- 笠 浩一郎, 于 海貝, 松原 茂樹: 同時通訳者の話速に影響を及ぼす要因の定量的分析, 通訳翻訳研究, No. 9, pp. 21-32 (2009).
- 遠山 仁美, 松原 茂樹: 英日同時通訳者発話におけるフィラーの出現と聴きやすさとの関係, 通訳研究(日本通訳学会誌), No. 7, pp. 39-49 (2007).
- 小野 貴博, 遠山 仁美, 松原 茂樹: 大規模音声コーパスを用いた日英・英日同時通訳における訳出遅延の比較分析, 通訳研究(日本通訳学会誌), No. 7, pp. 51-64 (2007).
- 遠山 仁美, 松原 茂樹: 訳出遅延時間と訳出開始タイミングに着目した同時通訳者の話者発話追従戦略に関する分析, 通訳研究(日本通訳学会誌), No. 6, pp. 113-128 (2006).
- Hitomi Tohyama, Shigeki Matsubara: Development of Web-based Teaching Material for Simultaneous Interpreting Learners using Bilingual Speech Corpus, Proceedings of World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA-2006), pp. 2906-2911 (2006).
- Hitomi Tohyama, Shigeki Matsubara: Collection of Simultaneous Interpreting Patterns by Using Bilingual Spoken Monologue Corpus, Proc. of 5th International Conference on Language Resources and Evaluation (LREC-2006), pp. 2564-2569 (2006).
- 遠山 仁美, 松原 茂樹: 同時通訳における聴きやすさとポーズの関係, ー同時通訳コーパスを用いた被験者実験による分析ー, 通訳研究(日本通訳学会誌), No.5, pp. 137-155 (2005).
- 大原 誠, 松原 茂樹, 笠 浩一郎, 河口 信夫, 稲垣 康善: 同時通訳を介した異言語間対話の時間的特徴 ー逐次通訳との比較に基づく対訳コーパスの分析ー, 通訳研究(日本通訳学会誌), No.3, pp. 34-52 (2003).
- 笠 浩一郎, 松原 茂樹, 稲垣 康善: 英日同時翻訳のための依存構造に基づく訳文生成手法", 電子情報通信学会論文誌, Vol. J92-D, No. 6, pp. 921-933 (2009).
- Koichiro Ryu, Shigeki Matsubara, Yasuyoshi Inagaki: Simultaneous English-Japanese Spoken Language Translation Based on Incremental Dependency Parsing and Transfer, Proceedings of COLING/ACL-2006, pp. 683-690 (2006).
- 大野誠寛, 松原茂樹: 文節間の依存・非依存を同定する漸進的係り受け解析, 電子情報通信学会論文誌, Vol. J98-D, No. 4, pp. 709-718 (2015).
- 大野 誠寛, 村田 匡輝, 松原 茂樹, 講演のリアルタイム字幕生成のための逐次的な改行挿入, 電気学会論文誌, Vol. 133-C, No. 2, pp. 418-426 (2013).
- 村田 匡輝, 大野 誠寛, 松原 茂樹: 読みやすい字幕生成のための講演テキストへの改行挿入, 電子情報通信学会論文誌, Vol. J92-D, No. 9, pp. 1621-1631 (2009).
- Tomohiro Ohno, Masaki Murata, Shigeki Matsubara: Linefeed Insertion into Japanese Spoken Monologue for Captioning, Proc. of ACL-IJCNLP-2009, pp. 531-539 (2009).