

2023 年（令和五年）度

第 37 回

日本音声学会全国大会予稿集

Proceedings of the 37th General Meeting of the PSJ

2023 年 9 月 16 日・17 日

September 16-17, 2023

日本音声学会

The Phonetic Society of Japan

フランス人学習者による自発英語発話における語彙アクセント自動測定

Sylvain Coulange^{1,2a,2b}, 加藤恒夫¹, Solange Rossato^{2a}, Monica Masperi^{2b}

¹ 同志社大学院理工学研究科音声言語処理機構研究室

^{2a} Univ. Grenoble Alpes, CNRS, Institute of Engineering, Grenoble Computer Science Laboratory (LIG), Grenoble, France

^{2b} Univ. Grenoble Alpes, Laboratory of Linguistics and Didactics of Foreign and Mother Tongues (LIDILEM), Grenoble, France

tsukato@mail.doshisha.ac.jp

{sylvain.coulange, solange.rossato, monica.masperi}@univ-grenoble-alpes.fr

1 はじめに

語彙アクセントは、単語内の強勢位置が変化する言語特徴であり、英語、ドイツ語やスペイン語などに現れる。これに対して、固定位置に強勢を持つフィンランド語（先頭の音節）、ポーランド語（末尾から2番目）やフランス語（末尾の音節）のような言語が対立する (Cutler & Jesse, 2021)。英語において、語彙アクセントは音節レベルの韻律およびスペクトル品質の変化として現れる。強勢のある音節は、通常、無強勢音節よりも長く、強く、高く、また F0 の変化が大きく、完全な母音品質を特徴とする (Cutler, 2015)。さらに、強勢音節は周囲の無強勢音節に影響を与え、短縮、弱体化させる (Tortel, 2021)。英語には様々な強勢、無強勢のレベルがあるが、本研究は第一強勢 (Primary stress) とその他の音節の対比に焦点を当てている。

語彙アクセントは単語分割と語義曖昧性解消の機能を果たす。内容語は一般に強勢を有し、機能語は弱化する傾向がある (Tortel, 2021)。語彙アクセントは派生形態論においても重要な役割を果たす。例えば、“PERson” と “perSONify” のように品詞に応じて語彙アクセントはしばしば変化し、“PHOTograph” と “phoTOgrapher” のように同一の品詞の単語を区別するのに役立つ。多くの名詞や形容詞は先頭の音節に強勢を持ち、動詞は先頭から2番目の音節に強勢を持つ。

第二言語習得の文脈では、学習者はしばしば母語の韻律規則の影響を受ける。たとえば、フランス語は単語末尾の音節に固定した強勢を有し、それ以外の母音には一貫した平板さがある。その結果、フランス語母語話者の英語はしばしば単語末尾に強勢位置を動かし、無強勢音節を弱体化しない傾向がある (Tortel & Hirst, 2010)。また、フランス語の強勢は語義曖昧性解消の機能を果たさないため、フランス語を母語とする英語学習者は強勢パターンの認識が限られており、特に句末延長や語彙アクセントを認識することが難しいとされる (Dupoux, Pallier, Sebastian, & Mehler, 1997)。このような状況を stress deafness と表現し、音節拍言語の母語話者は強勢拍言語の習得が難しいと指摘されている。さらに、リズムやイントネーションを意図的に変更するのは簡単なことではない。リズムやイントネーションは幼少期から個人に染みつき、性格や文化と密接に関連しているためである (Calbris & Montredon, 1975)。したがって、無強勢音節の弱体化や語彙アクセントの誤りは、聞き手の単語の識別を著しく妨げる可能性がある (Cutler, 2015)。また、フランス人英語学習者は、語彙アクセントの位置の改善、強勢音節と弱体化音節のコントラストの向上、無強勢音節の句末延長の回避、機能語の弱体化を重視すべきと文献 (Tortel, 2021) は述べている。

筆者らは、非母語話者の自発 L2 英語発話を書き起こし、語彙アクセントの位置と品質を特定する自動パイプライン^{*1}を開発した。多音節で構成される内容語の各音節を、基本周波数 (F0)、インテンシティおよび継続時間の3つの特徴量で評価し、単語単位の韻律形状を推定し、辞書の規範的な韻律形状と比較する。さらに、話者正規化を行った韻律特徴量から強勢音節と無強勢音節のコントラストを測り、各話者がどの韻律特徴量を用いて強勢を特徴付けているかを定量化する。

このパイプラインは、ヨーロッパ言語共通参照枠 (CEFR) の B1 と B2 レベルのフランス人 176 人による議論的な自発 L2 英語発話のコーパスに適用した。以下の節ではコーパス、パイプラインの詳細を示し、第4節では実験の結果について探求し、最終節で実験結果を考察する。

^{*1} パイプラインは以下で公開：<https://gricad-gitlab.univ-grenoble-alpes.fr/lidilem/plspp>

2 データ

データセットは、フランスにおける言語能力の国家試験である CLES^{*2} の対話スピーキング課題において録音された 176 名のフランス人英語学習者による L2 英語スピーチで構成されている。この課題は、2 名または 3 名の受験者が、例えば、電子タバコ、セキュリティカメラ、教室内のテクノロジーの使用など論争のあるトピックに関して 10 分間議論するロールプレイを含む。各受験者は 2 人の専門家により評価され、CEFR (Council of Europe, 2020) に基づきスピーキングレベル B2 以上、B1、B1 未満の判定が行われる。B1 未満の受験者はコーパスから除外した。その結果、176 人のうち 66 %が B2 レベル以上、34 %が B1 レベルであった。女性が 53 %、男性が 47 %であり、全受験生がフランス語を母語としていた。

3 方法論

自動処理のパイプラインは複数ステップで構成される。まず、Pyannote を用いてニューラル話者ダイアライゼーションを行う (Bredin & Laurent, 2021)。次に、WhisperX を用いて音声認識および単語レベルの強制アラインメントを行う (Bain, Huh, Han, & Zisserman, 2023)。その後、形態素解析に SpaCy を使用し、構成解析に Berkeley Neural Parser を使用する (Honnibal, Montani, Van Landeghem, & Boyd, 2020; Kitaev, Cao, & Klein, 2019)。録音音声は Pyannote の音声区間検出を用いて話者単位の連続音声セグメントに分割し、短い発話を除外するために、時間が 8 秒以上のセグメントのみを残す。これにより、合計 11 時間の連続音声コーパスが生成された。各話者の平均音声時間は 3 分 44 秒 (最小 32 秒, 最大 6 分 51 秒, 標準偏差 1 分 20 秒) であり、転記テキストには POS タグの注釈が施され、音声信号にアラインメントされる。

語彙アクセントの分析は、辞書の規範的な韻律形状を参照して強勢音節と無強勢音節のコントラストを定量化する。各音節は、話者正規化を行った 3 種類の特徴量、具体的には、音節核位置における F0、同位置におけるインテンシティ、単語境界もしくは隣接する音節核位置との中点を音節境界として得られる音節継続時間、で定量化される。音節核の位置は、インテンシティのピークに基づき音節核検出を行う Jong, Pacilly, & Heeren (2021) の Praat スクリプトを用いて決定する。音声以外のイベントの影響を最小限に抑えるために、事前に 300~3300Hz の帯域通過フィルタを適用する。転記テキスト中の各単語について、CMU 発音辞書^{*3} を参照して音節数を求め、検出した音節数と比較する。音節数が一致した単語のみ分析対象とする。“camera” や “chocolate” のように可能な音節数が複数ある場合、そのうち一つに一致していれば分析対象に含める。話者正規化は、話者ごとに絶対 F0、インテンシティ、および音節継続時間の分布を求め、各分布に基づきパーセンタイル値に変換した。これにより、話者ごとのパフォーマンスにおいて、各音節の特徴量の相対的な顕著性が求められる。最大のパーセンタイル値をもつ音節が観測された強勢音節、その他の音節は無強勢音節となる。

音節ごとに各特徴量のパーセンタイルを等重みで平均することで、語彙アクセントの総合的な表現を得る。語彙アクセントの位置は音節単位のバイナリ表現で分析する。ここでは、「O」が強勢音節、「o」がそれ以外の音節を表す。たとえば、“student” の規範的な韻律形状は「Oo」であり、先頭の音節に強勢が置かれ、末尾の音節は弱化する。また、“potential” の形状は「oOo」であり、二番目の音節に強勢がある。なお、本研究では第二強勢音節、無強勢音節、および弱化石節を区別せず、第一強勢音節の位置に焦点を当てる。

図 1 に本手法の出力例を示す。tier 1 と 2 に POS タグと転記テキスト、tier 3 に音節核の位置を示している。tier 4 に CMU 辞書が与える韻律形状、tier 5 に観測した 3 種類の特徴量から得た総合的な韻律形状、tier 6 に F0、インテンシティ、継続時間それぞれの観測値から得た韻律形状を示している。ここでは 2 値のアクセント表現のみ表示しているが、「o」と「O」のシンボルの背後にはパーセンタイル値

^{*2} <https://www.certification-cles.fr/english/>

^{*3} <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

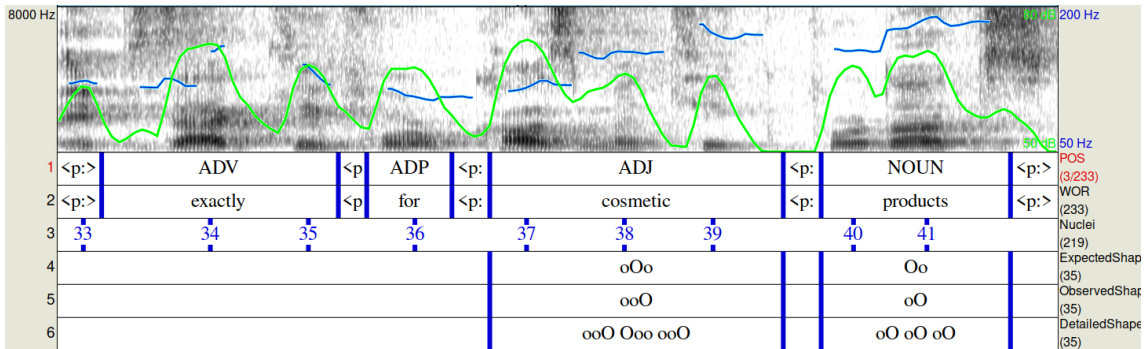


図 1: 自動処理パイプラインの出力例。品詞タグ (1), 転記テキスト (2), 音節核 (3), CMU 辞書の韻律形状 (4), 観測した 3 種類の特徴量から得た韻律形状 (5), F0, インテンシティ, 音節継続時間それぞれの観測値から得た韻律形状 (6)

が存在する。この例では、単語“exactly”の区間に 2 つの音節が検出されているが、CMU 辞書によれば 3 音節単語であるため、この単語は分析対象から外している。また、“cosmetic”と“products”の 2 つの単語はそれぞれ二番目と先頭の音節に強勢を有するが、音声からはどちらも末尾の音節に強勢が検出されている。

自動処理の精度について、著者のうちの一人が 28 の音声ファイルと 100 の対象単語を無作為に抽出し、手動で評価した。その結果、WhisperX の単語認識率は 92%、単語境界の正確な単語の割合は 95%、正しく音節核が検出された単語の割合は 87% であった。3 種類の特徴量から得た韻律形状と実際に知覚した強勢の一致率は 80% であった。強勢知覚は、90% の単語について確信をもって判断できた。ただし、主観的な作業であるため、今後複数の評価者で評価し、頑健性を確保する必要がある。

本研究の仮説としては、フランス人英語学習者による語彙アクセントは単語の規範的な韻律形状に関わらず末尾の音節に現れ、主に継続時間が長くなる傾向を示すと予想した。また、B2 レベル以上の学習者は、B1 レベル学習者よりも正確なアクセント位置を示し、強勢音節と無強勢音節の F0 とインテンシティのコントラストが大きくなると予想した。

4 結果

英語音声コーパス内に含まれる多音節の内容語 6,350 について強勢の位置と品質を調査した。これらの単語のうち、名詞は 57%、動詞は 18%、副詞は 13%、形容詞は 12% を占めた。2 音節からなる単語が多く (74%)、3 音節単語が 20%、4 音節単語が 5%、5 音節単語が 1% であった。B2 レベル以上と判定された学習者 (以下、B2 レベル学習者と呼ぶ) は、話速が速く、多音節の内容語を B1 レベルと判定された学習者 (以下、B1 レベル学習者と呼ぶ) よりも有意に多く含んだ。B2 レベル学習者の中央値が 47 単語に対して、B1 レベル学習者は 32 単語で、 $p < .001$ で有意差があった。一方、語数に対する多音節の内容語の割合に統計的に有意な差はなかった (B2 レベル学習者は 25%、B1 レベル学習者は 24%)。

最初に、韻律形状が CMU の韻律形状 (以下、参照形状と呼ぶ) と一致した単語の割合を調べた。全体で、参照形状と観測値から推定した韻律形状との一致率 (以下、一致率と呼ぶ) は 35.4% に留まった。話者ごとの一致率は 0% から 68.4% まで幅広く、中央値は 33.3% であった。

次に、B2 レベル学習者が B1 レベル学習者と比較して一致率が高いかどうか調べた。両グループの一致率の分布は分散が大きいですが、B2 レベル学習者の平均値が B1 レベル学習者の平均値よりも有意に大きかった。一致率の平均値は、36.0% と 29.6% であり、 $p < .0001$ で有意差があった。図 2 は、話者単位の対象語数と強勢位置の一致率の散布図である。一致率は、B1 レベル学習者のうち 2 人しか 50% を超えない一方、B2 レベル学習者のうち 26 人 (B2 グループの 22%) が 50% を超えた。

韻律形状の不一致は、単語を構成する音節数に比例して増加する傾向にある。2 音節単語で 62%、3

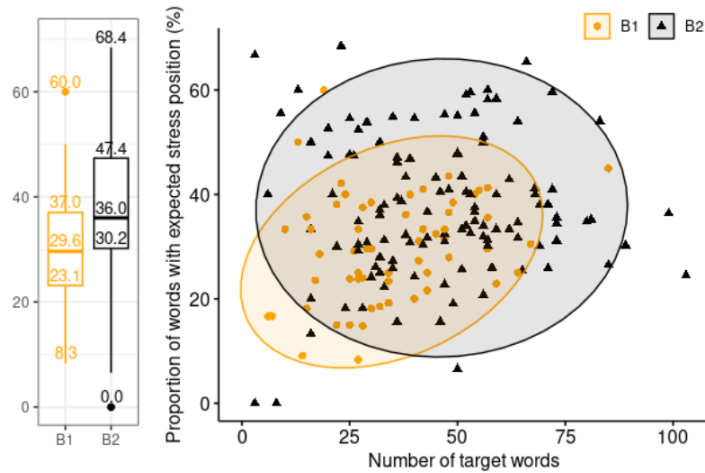


図 2: 話者単位の対象語数と強勢位置の一致率の散布図

音節単語で 70%, 4 音節単語で 79%, 5 音節単語で 81% であった。図3に、参照形状ごとに全話者から観測された韻律形状の割合を示す。特筆すべきは、2 音節単語のうち先頭の音節に強勢を有するものが 85% (4,029 語) を占めたが、実際にはそのうちの 31% (1,241 語) しか先頭の音節に強勢が置かれず、69% は末尾の音節に強勢が置かれていた。逆に、末尾の音節に強勢を有する「oO」の単語の大部分 (79%, 584 語) には正しい強勢が置かれていた。同様の傾向が 3 音節単語と 4 音節単語にも見られ、多くの単語は末尾の音節に強勢が置かれていた。

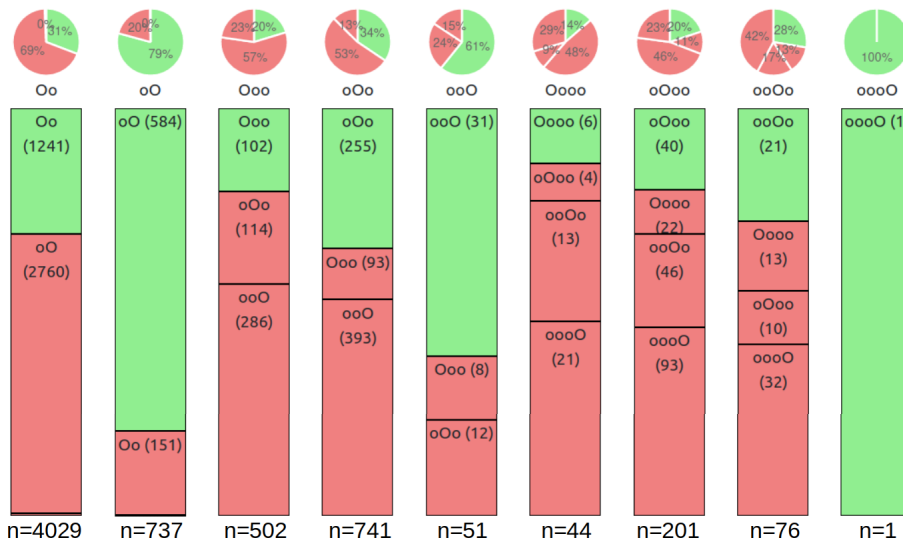


図 3: 規範的な語彙アクセントの韻律形状ごとのフランス人学習者による測定された韻律形状の割合 (度数)

B2 レベル学習者と B1 レベル学習者の比較では、B2 レベル学習者の一致率が平均的に B1 レベル学習者を上回る以外に、両グループ間に顕著な差は見られなかった。B2 レベル学習者と B1 レベル学習者の一致率の差は、「oOo」形状において 12 ポイント、「Oo」形状において 7 ポイント、「Ooo」形状において 6 ポイントであった。興味深いことに、「oO」形状における一致率は 5 ポイント低下した。これは過剰修正に起因している可能性がある。

また、韻律形状の一致の度数が多かった 20 の単語のうち 14 単語が、韻律形状の不一致の度数が多

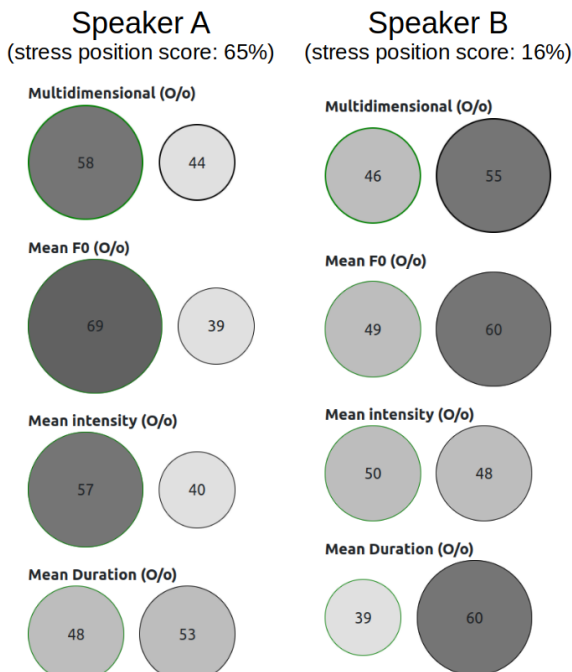


図 4: B2 レベルの学習者 2 名による強勢音節（左の円）とその他の音節（右の円）のパーセンタイル値。上から、3 種類の特徴量の平均による総合値、F0、インテンシティ、音節継続時間。

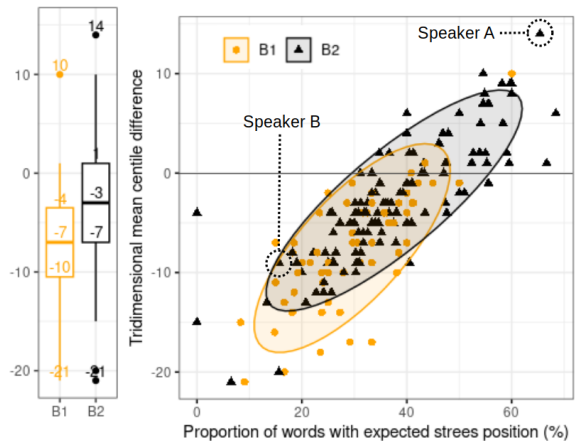
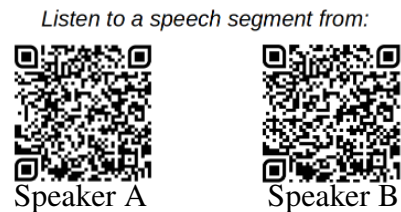


図 5: 話者単位の韻律形状の一致率と 3 種類の特徴量のパーセンタイル平均値の散布図

かった 20 の単語にも含まれた。頻度の高い単語ほど強勢位置が正しくなるわけではなかった。

図 4 は、B2 レベル学習者の 2 人が産出した単語の強勢音節とその他の音節の平均的なコントラストを示している。円の中に表示している数値はそれぞれパーセンタイル値の平均値であり、数値が大きいほど強調されていることを表す。Speaker A は 52 の対象単語のうち 65% の韻律形状が参照形状と一致していた。一方、Speaker B は 32 の対象単語のうち 16% しか一致していなかった。Speaker A の強勢音節はその他の音節よりも平均して、F0 は 30 ポイント、インテンシティは 17 ポイント大きく、継続時間はほとんど変わらなかった (-4 ポイント)。その結果、強勢音節とその他の音節の平均的なコントラストは 14 ポイントであった。一方、Speaker B は一致率が低かったため、負のコントラストを示した。強勢音節がその他の音節よりも平均して、継続時間は 21 ポイント、F0 は 11 ポイント小さく、インテンシティはほとんど変わらなかった (+3 ポイント)。この 2 つの傾向はそれぞれ、一致率の高い学習者群、一致率の低い学習者群に表れた。前者の群は、先ず強勢音節の F0、次にインテンシティを増加させ、継続時間はほとんど変化がない。一方、後者の群は、無強勢音節について F0 の上昇を伴って継続時間の増加が見られる一方、インテンシティはほとんど変化がない。

B2 レベル学習者と B1 レベル学習者の間で、強勢音節とその他の音節の平均音響コントラストの差は統計的に有意であった (B2 レベル学習者の中央値は-3、B1 レベル学習者の中央値は-7、 $p < .0001$)。また、両グループにおいて韻律形状の一致率と平均コントラストの間には強い相関があった ($R = .82$, $p < .0001$, 図5参照)。

5 考察

本研究では、B1 および B2 レベルのフランス人英語学習者 176 名について自発発話における語彙アクセントの一致率と、強勢音節と無強勢音節のコントラストを分析した。その結果、コーパスに含まれる 6,350 の多音節の内容語のうち、韻律形状が参照形状と一致した単語は 35.4% であり、話者により 0%

から 68.4% にわたる幅広いばらつきが観察された。B2 レベル学習者の一致率は 36% と、B1 レベル学習者の 29.6% よりも有意に高かった。予想どおり、語彙アクセントは単語ごとの韻律形状や音節数に関係なく、一貫して単語末尾の音節に集中する傾向が観察された。また、強勢の位置は音節の継続時間に大きく現れ、F0 とインテンシティのコントラストは韻律形状の一致率の高い学習者で大きくなった。

現時点での本研究の主な制約は、3 種類の特徴量に重み付けをせずに統合し、各特徴量の顕著性への寄与度を無視している点である。理論（例：Bolinger の Pitch theory of accent(1958)）を考慮すれば、強勢位置の決定には F0 パターンが主要な役割を果たすため、他の特徴量よりも F0 により多くの重みを割り当てるのが適切である可能性がある。ただし、英語のような強勢拍言語では、継続時間も重要な特徴として現れる (Grabe & Low, 2002)。インテンシティと継続時間を無視し、F0 のみで韻律形状決定すると、一致率は約 42% であった (B1 レベル学習者で 36%, B2 レベル学習者で 44%)。また、インテンシティのみで決定すると、この割合は 45% に上がった (B1 で 39%, B2 で 48%)。一方、継続時間だけで決定すると 30% に減少した (B1 および B2 の話者ともに)。

もう一つの制約は特徴量抽出に関するものである。現在の方法では音節核の位置で F0 を取得しているが、母音区間内の変動は考慮されていない。強勢音節は通常、基本周波数が大きく変化する傾向があるため、母音区間内の F0 変化の最小、最大、平均、方向など特徴量の追加を検討することが有益と考えられる。また、正確性を向上するには、音節区間内の子音、特に末尾の摩擦子音の延長が音節の継続時間に影響を与えている可能性があるため、音節ではなく母音の継続時間を考慮することも有益と考えられる。

韻律形状の精度を改善するために、音節核の検出と音素セグメンテーションを組み合わせることで母音区間の特徴量を抽出するとともに、F0 により大きい重みを割り当てることを検討する予定である。

■謝辞 本研究は JSPS 科研費 23H00648 の助成を受けたものである。
著者は IDEX による S.C. 留学助成金の提供を受けている。ここに謝意を表す。

■データとソースコードの提供状況

パイプラインは以下で公開：<https://gricad-gitlab.univ-grenoble-alpes.fr/lidilem/plspp>
音声データとメタデータは公開されている。詳細については、以下のアドレスにお問い合わせください：
coordination-nationale@certification-cles.fr

参考文献

- Bain, M., Huh, J., Han, T., & Zisserman, A. (2023) “Whisperx: Time-accurate speech transcription of long-form audio.” *Interspeech*.
- Bolinger, D. L. (1958) “A theory of pitch accent in english.” *WORD*, 14, 109-149.
- Bredin, H., & Laurent, A. (2021) “End-to-end speaker segmentation for overlap-aware resegmentation.” In *Interspeech*.
- Calbris, G., & Montredon, J. (1975) *Approche rythmique, intonative et expressive du français langue étrangère*. CLES International.
- Council of Europe (2020) *Common european framework of reference for languages*. Strasbourg, France: Council of Europe.
- Cutler, A. (2015) “Lexical stress in english pronunciation.” In *The handbook of english pronunciation*. (pp. 106–124). Hoboken, NJ: John Wiley & Sons, Inc.
- Cutler, A., & Jesse, A. (2021) “Word stress in speech perception.” John Wiley & Sons, Ltd.
- Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997) 「A Destressing “Deafness” in French?」 『Journal of Memory and Language』 36: 3, 406-421.
- Grabe, E., & Low, E. L. (2002) “Durational variability in speech and the rhythm class hypothesis.” *Papers in Laboratory Phonology*, 7, 515-546.
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020) *spacy: Industrial-strength natural language processing in python*.
- Jong, N. H. de, Pacilly, J., & Heeren, W. (2021) “Praat scripts to measure speed fluency and breakdown fluency in speech automatically.” *Assessment in Education: Principles, Policy & Practice*, 28, 456-476.
- Kitaev, N., Cao, S., & Klein, D. (2019) “Multilingual constituency parsing with self-attention and pre-training.” In *ACL* (pp. 3499–3505). Florence, Italy.
- Tortel, A. (2021) “Le rythme en anglais oral : considérations théoriques et illustrations sur corpus.” *Recherche et pratiques pédagogiques en langues - Cahiers de l’APLIUT*.
- Tortel, A., & Hirst, D. (2010) “Rhythm metrics and the production of English L1/L2.” *Speech Prosody 2010*, Paper 959.