On Entrainment in Semi-Spontaneous Multilingual Parliamentary Speech

Jennifer Jane Ries[1,*,**], *Debasmita Bhattacharya*[1,*], *Julia Hirschberg*[1]

[1] Department of Computer Science, Columbia University, USA

jr3954@columbia.edu, debasmita.b@cs.columbia.edu, julia@cs.columbia.edu

Abstract

Entrainment has predominantly been studied in monolingual instances of either spontaneous or task-oriented speech, leaving open questions of how speakers adapt to their interlocutors in mixed-spontaneity settings where dialogue can also span multiple languages. Our work studies entrainment in the semi-spontaneous context of the Hansard French-English parliamentary proceedings. We examine previously unexplored acoustic-prosodic features of Hansard's spoken exchanges alongside their semantic aspects at multiple levels of temporal granularity, finding that speakers adapt across inter-related dimensions of entrainment and cross-lingual language settings in accordance with primacy and recency effects. Our findings give rise to novel insights into potential planning and priming mechanisms in an understudied context, motivating further work on the entrainment dynamics of linguistically diverse speech settings.

Index Terms: entrainment, multilingual, conversational dynamics, speech analysis, multimodal

1. Motivation

Speakers communicate in accordance with a range of conversational goals, from signaling social identity and fostering rapport to ensuring mutual comprehension and creating a positive impression. One way to achieve such goals is by *entraining* to one's interlocutors. *Entrainment* – also referred to as *linguistic accommodation*, *alignment*, or *coordination* – occurs when speakers subconsciously adapt aspects of their communication style to match that of their conversation partner. Entrainment can take place over multiple dimensions and modalities of language production including diction and syntax [1], speaking rate, voice quality, and pause frequency [2, 3, 4], and even facial expression and gesture [5, 6, 7]. Much prior work has focused on studying entrainment in *speech*; early studies defined metrics and statistical frameworks needed to examine this linguistic phenomenon with respect to lexical, acoustic, prosodic, and semantic features of spontaneous and task-oriented conversation [8, 3, 9]. Extensions of these studies showed that entrainment is positively correlated with greater dialogue success and desirable social outcomes [10, 11] and may be modulated by paralinguistic factors such as interlocutor gender and power differentials [10, 12, 1]. More recent work on entrainment in speech has incorporated machine learning approaches [13, 14] and explored interactions between its various linguistic dimensions [15, 16].

Despite the breadth of existing work on spoken entrainment, prior studies have predominantly examined monolingual instances of either spontaneous open-domain speech or specialized task-oriented dialogue in extended turn-taking [3, 4]. This has left open questions of how entrainment occurs in settings where dialogue involves *variable* levels of spontaneity, such as during legal proceedings, and is uttered in *multilingual* contexts. A few prior studies considered semi-spontaneous speech settings but focused entirely on U.S. Supreme Court proceedings [17, 18, 19], limiting their generalizability to languages other than English as spoken in the United States. Separately, while some authors have explored multilingual lexical representations [20, 21, 22, 23], few have studied spoken multilingual entrainment [24]. The combination of these factors has left wide scope for exploring how entrainment takes place in *semi-spontaneous and multilingual speech contexts*, especially at the level of single-turn exchanges as in [25, 26, 27, 28]. To address this gap, we ask the following **RQ:** How does a speaker adapt their communication style with respect to the interlocutor within a dialogue that involves variable levels of spontaneity and may require communicating across multiple languages?

To answer this **RQ**, we study the Hansard corpus [29] of Canadian French-English parliamentary speech and examine downstream entrainment dynamics between long prepared monologues and shorter spontaneous responses. We find that **1)** speakers entrain monolingually *and* cross-lingually in accordance with primacy and recency effects [30]; **2)** these patterns of entrainment are present in multiple inter-related linguistic dimensions; and **3)** these effects are influenced by both the presence and the identity of the languages involved in dialogue. Overall, consistent with entrainment as a *universal* dialogue behavior, the patterns we find in this linguistically diverse spoken context mirror many of those originally seen in monolingual speech settings with fixed levels of spontaneity. Additional temporal effects and interdependence between dimensions of entrainment that appear notably in multilingual contextual conditions indicate important ways in which mixed-spontaneity cross-lingual dialogue may be discursively *unique*. These in turn have implications for potential planning and priming mechanisms underpinning conversational dynamics in both monolingual and multilingual single-turn exchanges, particularly regarding the influence of speakers' memory of prolonged speech. Our work is the first to study entrainment within the Canadian Hansard corpus, contributing new insight into the fine-grained and interactive entrainment dynamics of monolingual and multilingual dialogue in an underexplored speech setting.

2. Data

We study a subset of the Canadian Hansard corpus of French-English legislative discourse [29]. Hansard is continually updated with the most recent parliamentary debates, with publicly-available audio recordings available from September 2006 on. We focus on Hansard proceedings with available recordings

*Equal contribution.

**Corresponding author.

from the first session of the 39th Parliament, taking place between 2006 and 2007.[1] This part of Hansard involves 175 chamber meetings of the Canadian parliament and 317 speakers (20 'Presiding Officer', 102 'Honourable', and 195 untitled speakers). Transcribed and translated utterances include prepared speeches and legislative propositions and spontaneous questions and discussion. The data contain monolingual English, monolingual French, and code-switched French-English utterances, each annotated with utterance-level language ID tags.

Prior studies leveraged Hansard's parallel text format for machine translation applications [32]; our work offers a new perspective on the dataset's previously unexplored acoustic-prosodic aspects. From our subset of Hansard, we exclude turns containing intrasentential code-switching and randomly sample 40 dyadic mixed-spontaneity exchanges from the remaining monolingual data. Each exchange comprises a prepared monologue (10-20 minutes) followed by a spontaneous question (1-2 minutes), totaling over 12 hours of recorded speech and 7,500 utterances. Each monologue-question instance takes one of four language settings: *en-en*, *fr-fr*, *en-fr*, or *fr-en*. We ensure balanced subsets of each language setting and remove power differentials between interlocutors by excluding speakers with special titles, mitigating a known confound of entrainment [12, 1].

3. Method

Our goal is to evaluate how a questioner's speech entrains to that of the monologue to which they responded.

Feature extraction. For each exchange, we extract the 88 *acoustic-prosodic* features in eGeMAPSv02 [33] (spanning F0, energy, spectral, cepstral, voice quality, and duration measures) to vectorize both the monologue and the question; we choose this set for its compact yet cross-lingually representative feature space [34, 35]. Guided by [36], we obtain an acoustic-prosodic baseline for the questioner using the features of a monologue uttered by that speaker in the same language elsewhere in the corpus; we use it to normalize the monologue and question vectors in the exchange. We also use each exchange's transcript to create 1152-dimensional *semantic* representations of the monologue and question using the multilingual encoder RemBERT,[2] chosen for its cross-lingually robust embeddings [37].

Quantifying entrainment. For both acoustic-prosodic and semantic representations, we follow [38, 39, 40] and use cosine similarity to measure global entrainment between normalized monologue and question vectors per exchange. We first obtain null mean entrainment values to capture that the expected global entrainment between a monologue and non-corresponding question is near zero [41]. For both *acoustic-prosodic* and *semantic* feature sets, we calculate these null values for *en-en*, *fr-fr*, *en-fr*, and *fr-en* using the cosine similarity between 5000 random pairings of monologues and questions from each language setting; we derive p-values by comparing observed scores against these null values. To avoid assuming uniform entrainment in each dimension across monologues, we study fine-grained dynamics by dividing each monologue into its beginning, middle, and end and repeat the above calculations to compare each third to the associated question. We obtain null mean entrainment for each exchange from the mean cosine similarity across monologue segments with the subsequent question. We apply further Pearson correlation testing and linear mixed-effects modeling using `statsmodels 0.14.6`.

Explaining entrainment. We qualitatively determine which features contribute the most to entrainment and calculate the proportion of entrainment each accounts for. We first identify the top contributor to each dimension of global entrainment per exchange, resulting in the top-10 global contributors per language setting. We quantitatively verify their influence by ablating them and recalculating global entrainment, as above. We repeat our qualitative steps at the temporal level by determining the acoustic-prosodic and semantic features with the greatest contributions to entrainment between each of the three monologue segments and the question within a given exchange.

4. Results

4.1. Global entrainment mirrors other speech settings

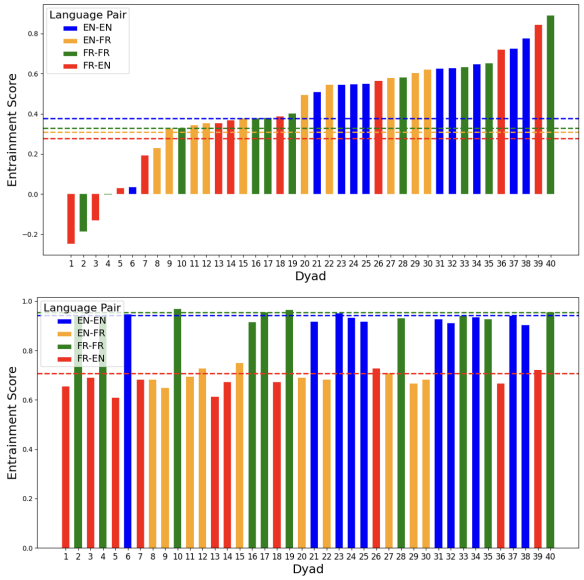

Figure 1: *Global acoustic-prosodic (top) and semantic (bottom) entrainment per exchange. Dotted lines indicate the relevant statistical significance threshold for each language setting.*

We first examine global entrainment of questions to monologues across the four language settings and both dimensions of entrainment. We find widespread *acoustic-prosodic* entrainment (Figure 1; top): 57.5% of exchanges exhibit significant above-null entrainment ($p < 0.05$) and 20% more approach significance ($p < 0.1$). We find similar, but weaker, signs of significantly non-null *semantic* entrainment on 17.5% of exchanges, with 10% more approaching significance (Figure 1; bottom); this validates our method, since questions could in principle be unrelated to their preceding monologues. The reduction in entrainment compared to acoustic-prosodic features is expected: since monologues are longer and cover more semantic topics, a question cannot realistically semantically align above inherently high baseline similarity with a monologue in its entirety.

The observed dynamics on each dimension of entrainment generally mirror those from fixed-spontaneity monolingual settings [3, 16], suggesting that patterns of entrainment from these spoken contexts may well apply to semi-spontaneous, multilingual settings. While some exchanges entrain more weakly than others, this remains aligned with prior work, as it is rare for en-

[1] Although audio data was retained on the Canadian House of Commons' website, it was not consolidated with the remainder of the corpus. To match transcript data with its audio, we manually downloaded and aligned audio for relevant exchanges using the ParlVU interface [31].

[2] `https://huggingface.co/google/rembert`

tire corpora to exhibit uniformly strong entrainment across data points. In contrast, while prior work disagrees on the extent to which distinct dimensions of entrainment coincide [15, 16], we find initial indications of *overlapping* acoustic-prosodic and semantic entrainment in highly-entraining exchanges in Hansard.

We investigate a possible relationship between dimensions of significant global entrainment and whether it differs by language setting. Acoustic-prosodic and semantic entrainment are *negatively* correlated in *monolingual* exchanges ($r \approx -0.3$), and moderately so in *en-en* ones ($r \approx -0.5$). In contrast, these are *positively* correlated in *cross-lingual* exchanges ($r \approx 0.3$), and moderately so in *fr-en* ones ($r \approx 0.4$). While none achieve statistical significance, the contrast between monolingual and cross-lingual correlation signs could indicate a potential high-level coupling between the two dimensions of entrainment that varies by language setting and warrants further investigation.

4.2. Temporal entrainment exhibits primacy and recency

Since global measures could obscure finer-grained properties, we next quantify how entrainment in questions develops across the monologue timescale. On *acoustic-prosodic* features, questioners entrain more strongly to the beginning and end of their corresponding monologue, relative to the monologue's middle. These patterns reflect primacy and recency effects of memory, where people best recall the first and last items in a series while forgetting the middle [30], and appear significantly across language settings in 82.5% of exchanges. Comparing language settings, primacy and recency are the most pronounced for questions in English (*en-en* and *fr-en* exchanges), suggesting that the language used in entraining speech, rather than entrained-upon speech, may characterize temporal alignment patterns.

We find similar primacy and recency of *semantic* entrainment in 67.5% of exchanges, with greater primacy than recency. This may be due to deeper attention to the part of a monologue where key arguments are typically introduced, as is required to hold these in memory long enough to align appropriately. These semantic effects are most pronounced for questions in French, complementing the English-specific patterns of the strongest temporal entrainment over acoustic-prosodic features. These temporal trends across dimensions of entrainment broadly align with our global observations and reveal continued mirroring of acoustic-prosodic and semantic entrainment, even at a more granular level of temporal analysis.

Given subtle differences in our temporal findings so far, we begin a deeper study of how dimensions of entrainment interact: we first hypothesize that generally, *semantic* entrainment has stronger *primacy* effects while *acoustic-prosodic* entrainment has stronger *recency* effects. To test this, we fit a mixed effects model,[3] calculating estimated marginal means and planned contrasts on the interaction between entrainment dimensions (semantic vs. acoustic-prosodic) and monologue temporal segments (initial vs. final). We find no evidence of general semantic primacy, nor general acoustic-prosodic recency ($\beta = 0.003$ and $p > 0.05$ for both), and observe null interaction between entrainment dimensions and segment locations ($\beta = 0.005$, $p > 0.05$), indicating that temporal profiles of semantic and acoustic-prosodic entrainment do not generally differ this way.

Despite the null interaction between acoustic-prosodic and semantic temporal entrainment, we revisit their opposing global correlations in different language settings, and now hypothesize that temporal asymmetries between dimensions of entrainment only surface under specific cross-lingual constraints. We test this using a mixed effects model (Table 1) and assess whether a question's semantic entrainment with its monologue's initial segment predicts its acoustic-prosodic entrainment with that monologue's final segment in cross-lingual exchanges.

Table 1: *Summary of mixed effects model: FinalAcoustic ∼ InitialSemantic * LanguageSetting. FinalAcoustic and InitialSemantic (IS) represent the entrainment score between a question and the final and initial segments of its paired monologue, respectively. LanguageSetting (LS) categorically encodes whether a given exchange is monolingual (coded 0) or cross-lingual (coded 1). Bolded features are significant.*

Factor	coef	std err	z	P>|z|
Intercept	2.276	1.254	1.816	0.069
IS	-1.898	1.344	-1.412	0.158
LS	**-2.451**	0.718	-3.414	**0.001**
IS * LS	**2.812**	0.578	4.862	**0.000**

We find that the relationship between initial semantic and final acoustic-prosodic entrainment is significantly more positive in cross-lingual exchanges than in monolingual ones (Table 1; interaction effect). While *en-fr* and *fr-en* exchanges exhibit lower entrainment overall (Figure 1), within these exchanges, a questioner who entrains semantically to a monologue's opening is significantly more likely to also align acoustically with its close. In monolingual exchanges, initial semantic entrainment has the opposite relationship with final acoustic-prosodic entrainment, exhibiting a weak trade-off between the two dimensions. This contextual *interdependence* between dimensions of entrainment may reflect alignment that occurs in stages; the strong association between initial semantic and final acoustic-prosodic entrainment suggests that speakers first align to monologues semantically to plan a question's *content*, and only scaffold its acoustic-prosodic *form* immediately before producing it. This particularly applies to speakers who may face greater cognitive constraints for achieving *cross-lingual* entrainment compared to monolingual cases. These language-setting-dependent patterns are consistent with those at the global level, enhancing our finer-grained understanding of interacting acoustic-prosodic and semantic entrainment in Hansard and pointing to a two-stage planning mechanism unique to cross-lingual dialogue.

4.3. A minority of features drives a majority of entrainment

We now interpret the factors driving such global and temporal entrainment. While we note some *acoustic-prosodic* patterns that hold across language settings – e.g., all F0 aspects (mean, range, etc.) recur as high-impact features in all cases – the top contributors to global entrainment seem to depend on the language setting of a given exchange and the specific languages present. In exchanges involving English, Mel-frequency cepstral coefficient (MFCC) features recur as top contributors to global entrainment, accounting for up to 26% acoustic-prosodically. In exchanges involving French, loudness features, e.g., its standard deviation, consistently contribute the most, accounting for up to 8% of global entrainment. Also, individual acoustic-prosodic features, rather than cohesive feature groups, contribute more to entrainment in cross-lingual than in monolingual exchanges, as the mean top feature contribution is 11% vs. 6% globally. This reveals another difference in entrainment style over different language settings. Temporal analyses of feature contributions reinforce global patterns, as the highest-

[3]Entrainment ∼ FeatureType * SegmentLocation + LanguageSetting, with a random intercept for each exchange.

contributing global acoustic-prosodic features exhibit similar importance across temporal segments under matching language setting constraints. Overall, percentage contributions are disproportionately high relative to acoustic-prosodic feature count.

Semantic entrainment exhibits similar patterns, both across and within language settings. The highest-impact embedding index is consistent over language settings and global and temporal comparisons, activating tokens dealing with public policy and legislative decisions such as 'illégaux', 'l'administration', and 'non-partisan', i.e., the core topics in Hansard. Other high-impact global and temporal contributors include an index activated by numerical concepts ('third', '5,000', 'milliers'), which drives 3-6% of temporal entrainment across language settings, and an index activating direct address-related tokens ('Monsieur', 'Honourable', 'Speaker') that accounts for 2-4% of temporal entrainment across language settings. Percentage contributions here are disproportionately even higher given the larger semantic feature space. Strikingly, although all exchanges across language settings share the same highest-impact index, this too exhibits significant language-specific patterns: its contribution to global entrainment is higher in *en-en* exchanges (25% of such entrainment) than in *fr-fr* ones (16%), and is intermediate for cross-lingual exchanges (19% of *en-fr* and *fr-en* entrainment). This reflects a relatively stronger influence on semantic entrainment involving English, which also replicates for the following two highest-impact semantic factors.

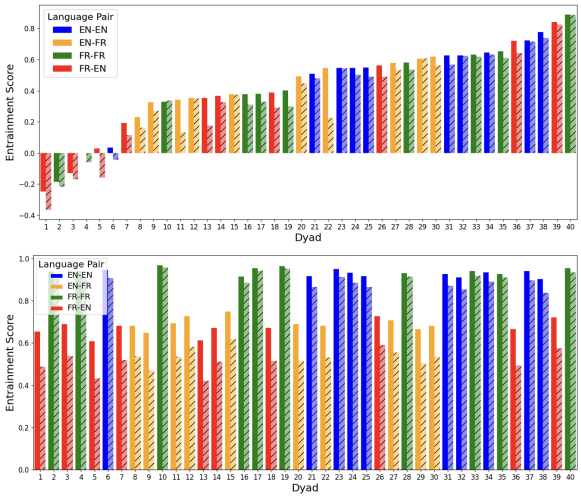

Figure 2: *Global acoustic-prosodic (top) and semantic (bottom) entrainment per exchange. Hatched bars represent entrainment following ablation of the highest-impact features in each case.*

Figure 2 demonstrates the impact of the primary drivers of global entrainment in each dimension: in their absence, entrainment decreases by up to 75% acoustic-prosodically and up to 24% semantically. These correspond to mean reductions of 14% and 12% across each dimension of entrainment, respectively; the latter is strongly significant on semantic features. We also observe that reductions in entrainment from removing the highest-impact features are consistently greater in cross-lingual exchanges. Absent the top contributors, acoustic-prosodic entrainment decreases by 29% and 18% in *fr-en* and *en-fr* exchanges, compared to 6% and 10% in *en-en* and *fr-fr* ones. Similarly, the absence of top contributors reduces semantic entrainment by 24% and 22% in *fr-en* and *en-fr* exchanges, compared to 5% and 2% in *en-en* and *fr-fr* ones. So, the strongest drivers of entrainment have greater impact on cross-lingual exchanges. Overall, we identify the features determining *how* entrainment occurs in each dimension in Hansard, showing their consistency over both global and temporal levels of qualitative analysis, and with respect to revealing language-setting-dependent patterns.

5. Discussion

We explore an understudied semi-spontaneous and multilingual speech setting, uncovering insights into its unique entrainment dynamics. Our global entrainment results mirror prior work on fixed-spontaneity monolingual contexts. Our temporal approach reveals new primacy and recency effects with contextual interdependency between dimensions of finer-grained entrainment in this setting; this suggests selective priming over time by different dimensions of entrainment, shaping the entrained production of higher-spontaneity responses to prepared speech. Such two-staged entrainment implies the priming of semantic content is a slower and more deliberate process that draws on earlier and more memorable speech, relative to faster and more reflexive acoustic-prosodic priming that may be triggered by recent auditory exposure. Such inter-relatedness in the underlying planning mechanism of entrainment has not been clearly shown in speech settings with fixed spontaneity, where authors disagree on the extent to which dimensions of global entrainment interact [15, 16]. Our temporal approach instead addresses the fine-grained, multidimensional aspects of prolonged interlocutor speech that a speaker must recall well enough to entrain, proposing a cognitive perspective that complements prior work on both the social motivators of entrainment [10] and its connections to memory and priming [28].

We also offer a fresh multilingual perspective. First, our proposed multidimensional planning mechanism emerges only in cross-lingual exchanges. We speculate that monolingual dialogue offers an advantage where priming can occur *simultaneously* across entrainment dimensions since the questioner expresses both content and form in the same language as the interlocutor. In cross-lingual dialogue, in contrast, the questioner reformulates content across languages and may be unable to mimic exact acoustic-prosodic aspects when making that language switch, which could trigger the integrated, but *staggered*, planning of content and form. Separately, while we find broad agreement across exchanges that supports a *universal* quality of entrainment [24], we also discover language-specific dynamics, some unique to cross-lingual dialogue. Certain temporal and feature-level effects in monolingual exchanges are language-dependent. And, though typically weaker cross-lingually, entrainment in such exchanges relies *more* on the most influential acoustic-prosodic and semantic drivers. This supports a potential compensatory strategy used by speakers facing more barriers to entrainment than those engaged in monolingual dialogue. These patterns together indicate linguistic constraints that strikingly interact with monolingual and cross-lingual entrainment.

Based on our findings, we argue that entrainment is a universal dialogue behavior and that studying it in diverse discourse contexts enables more nuanced and comprehensive insights into spoken conversational dynamics. Our work is a first step, inviting new research on cross-lingual entraining responses among bilingual interpreters, code-switchers, and language learners. We focus on downstream entrainment in single-turn mixed-spontaneity exchanges, a design choice that permits careful statistical control and yields multiple significant results. Fruitful extensions could apply our approach to multi-turn semi-spontaneous conversations, including other languages or the use of alternative feature sets, as well as to larger corpora.

6. Acknowledgments

We thank Nicholas Deas for helpful discussions and feedback. DB is supported by the National Science Foundation under Grant IIS 2418307.

7. Use of Generative AI Disclosure

The authors did not use generative AI tools in this work.

8. References

[1] C. Danescu-Niculescu-Mizil, M. Gamon, and S. Dumais, "Mark my words! Linguistic Style Accommodation in Social Media," in *Proceedings of the 20th International Conference on World Wide Web*. ACM, March 2011.

[2] H. Giles, N. Coupland, and J. Coupland, *Accommodation theory: Communication, context, and consequence*, ser. Studies in Emotion and Social Interaction. Cambridge University Press, 1991, p. 1–68.

[3] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," in *Proc. Interspeech 2011*, 2011, pp. 3081–3084.

[4] R. Chen, S. Kim, A. Papangelis, J. Hirschberg, Y. Liu, and D. Hakkani-Tür, "Identifying Entrainment in Task-Oriented Conversations," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[5] R. E. Maurer and J. H. Tindall, "Effect of postural congruence on client's perception of counselor empathy," *Journal of Counseling Psychology*, vol. 30(2), pp. 158–163, 1983.

[6] J. K. Burgoon, D. B. Buller, J. L. Hale, and M. A. de Turck, "Relational Messages Associated with Nonverbal Behaviors," *Human Communication Research*, vol. 10, no. 3, pp. 351–378, 03 2006.

[7] T. L. Chartrand and J. A. Bargh, "The chameleon effect: the perception-behavior link and social interaction." *Journal of personality and social psychology*, vol. 76 6, pp. 893–910, 1999.

[8] A. Nenkova, A. Gravano, and J. Hirschberg, "High Frequency Word Entrainment in Spoken Dialogue," in *Proceedings of ACL-08: HLT, Short Papers*. Columbus, Ohio: Association for Computational Linguistics, Jun. 2008, pp. 169–172.

[9] A. Gravano, Š. Beňuš, R. Levitan, and J. Hirschberg, "Three ToBI-based measures of prosodic entrainment and their correlations with speaker engagement," in *2014 IEEE Spoken Language Technology Workshop (SLT)*, 2014, pp. 578–583.

[10] R. Levitan, A. Gravano, L. Willson, Š. Beňuš, J. Hirschberg, and A. Nenkova, "Acoustic-Prosodic Entrainment and Social Behavior," in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Montréal, Canada: Association for Computational Linguistics, Jun. 2012, pp. 11–19.

[11] R. H. Gálvez, L. Gauder, J. Luque, and A. Gravano, "A unifying framework for modeling acoustic/prosodic entrainment: definition and evaluation on two large corpora," in *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 1st virtual meeting: Association for Computational Linguistics, Jul. 2020, pp. 215–224.

[12] J. S. Pardo, "On phonetic convergence during conversational interaction," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, 04 2006.

[13] M. Nasir, B. Baucom, S. Narayanan, and P. Georgiou, "Towards an Unsupervised Entrainment Distance in Conversational Speech Using Deep Neural Networks," in *Interspeech 2018*, 2018, pp. 3423–3427.

[14] M. McNeill and R. Levitan, "Autoregressive cross-interlocutor attention scores meaningfully capture conversational dynamics," in *Interspeech 2024*, 2024, pp. 2945–2949.

[15] Y. Liu, A. Li, J. Dang, and D. Zhou, "Semantic and Acoustic-Prosodic Entrainment of Dialogues in Service Scenarios," in *Companion Publication of the 2021 International Conference on Multimodal Interaction*, ser. ICMI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 71–74.

[16] J. Kejriwal and Š. Beňuš, "Relationship between auditory and semantic entrainment using Deep Neural Networks (DNN)," in *Interspeech 2023*, 2023, pp. 2623–2627.

[17] Š. Beňuš, R. Levitan, and J. Hirschberg, "Entrainment in spontaneous speech: The case of filled pauses in Supreme Court hearings," in *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*, 2012, pp. 793–797.

[18] Š. Beňuš, A. Gravano, R. Levitan, S. I. Levitan, L. Willson, and J. Hirschberg, "Entrainment, dominance and alliance in Supreme Court hearings," *Know.-Based Syst.*, vol. 71, no. 1, p. 3–14, Nov. 2014.

[19] C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, and J. Kleinberg, "Echoes of power: language effects and power differences in social interaction," in *Proceedings of the 21st International Conference on World Wide Web*, ser. WWW '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 699–708.

[20] G. J. Kootstra, J. G. van Hell, and T. Dijkstra, "Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue," *Journal of Memory and Language*, vol. 63, no. 2, pp. 210–231, 2010.

[21] E. Ahn, C. Jimenez, Y. Tsvetkov, and A. W. Black, "What Code-Switching Strategies are Effective in Dialog Systems?" in *Proceedings of the Society for Computation in Linguistics 2020*. New York, New York: Association for Computational Linguistics, Jan. 2020, pp. 254–264.

[22] T. Parekh, E. Ahn, Y. Tsvetkov, and A. W. Black, "Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues," in *Proceedings of the 24th Conference on Computational Natural Language Learning*. Online: Association for Computational Linguistics, Nov. 2020, pp. 565–577.

[23] A. Bawa, M. Choudhury, and K. Bali, "Accommodation of Conversational Code-Choice," in *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 82–91.

[24] D. Bhattacharya, S. Ding, A. Nguyen, and J. Hirschberg, "Measuring Entrainment in Spontaneous Code-switched Speech," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 2865–2876.

[25] I. Ivanova, W. S. Horton, B. Swets, D. Kleinman, and V. S. Ferreira, "Structural alignment in dialogue and monologue (and what attention may have to do with it)," *Journal of Memory and Language*, vol. 110, pp. 1–18, 2020.

[26] C. J. Wynn, S. A. Borrie, and T. P. Sellers, "Speech Rate Entrainment in Children and Adults With and Without Autism Spectrum Disorder," *American Journal of Speech-Language Pathology*, vol. 27, no. 3, pp. 965–974, 2018.

[27] S. A. Borrie and J. M. Liss, "Rhythm as a Coordinating Device: Entrainment with Disordered Speech," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 3, pp. 815–824, 2014.

[28] K. Weatherholtz, K. Campbell-Kibler, and T. F. Jaeger, "Socially-mediated syntactic alignment," *Language Variation and Change*, vol. 26, no. 3, p. 387–420, 2014.

[29] Canadian House of Commons. Debates (Hansard) - Parliament of Canada. [Online]. Available: https://www.ourcommons.ca/documentviewer/en/39-1/house/hansard-index

[30] J. Jahnke, "Primacy and recency effects in serial-position curves of immediate recall," *Journal of Experimental Psychology*, vol. 70, pp. 130–132, 07 1965.

[31] Parliament of Canada. ParlVU. [Online]. Available: https://parlvu.parl.gc.ca/

[32] P. F. Brown, J. Cocke, S. A. Della Pietra, V. J. Della Pietra, F. Jelinek, J. D. Lafferty, R. L. Mercer, and P. S. Roossin, "A Statistical Approach to Machine Translation," *Computational Linguistics*, vol. 16, no. 2, pp. 79–85, 1990.

[33] M. Yang, J. Konan, D. Bick, A. Kumar, S. Watanabe, and B. Raj, "Improving Speech Enhancement through Fine-Grained Speech Characteristics," in *Interspeech 2022*, 2022, pp. 2953–2957.

[34] D. Porjazovski, Y. Getman, T. Grósz, and M. Kurimo, "Advancing Audio Emotion and Intent Recognition with Large Pre-Trained Models and Bayesian Inference," in *Proceedings of the 31st ACM International Conference on Multimedia*, ser. MM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 9477–9481.

[35] A. S. G. Choi, J.-s. Kim, S.-h. Kim, M. S. Back, and S. Cho, "Crosslinguistic Acoustic Feature-based Dementia Classification Using Advanced Learning Architectures," in *Proceedings of the Fifth Workshop on Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric/developmental impairments @LREC-COLING 2024*, D. Kokkinakis, K. C. Fraser, C. K. Themistocleous, K. L. Fors, A. Tsanas, and F. Ohman, Eds. Torino, Italia: ELRA and ICCL, May 2024, pp. 95–100.

[36] Z. Yuan, D. de Jong, Š. Beňuš, N. Nguyen, R. Feng, R. Sabo, L. Fadiga, and A. D'Ausilio, "ART: The Alternating Reading Task Corpus for Speech Entrainment and Imitation," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, and N. Xue, Eds. Torino, Italia: ELRA and ICCL, May 2024, pp. 1548–1562.

[37] H. W. Chung, T. Févry, H. Tsai, M. Johnson, and S. Ruder, "Rethinking Embedding Coupling in Pre-trained Language Models," in *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

[38] J. Kejriwal, Š. Beňuš, and L. M. Rojas-Barahona, "Entrainment detection using DNN," *Computer Speech Language*, vol. 99, p. 101930, 2026.

[39] C. Dideriksen, M. H. Christiansen, M. Dingemanse, M. Højmark-Bertelsen, C. Johansson, K. Tylén, and R. Fusaroli, "Language-Specific Constraints on Conversation: Evidence from Danish and Norwegian," *Cognitive Science*, vol. 47, no. 11, p. e13387, 2023.

[40] N. D. Duran, A. Paxton, and R. Fusaroli, "ALIGN: Analyzing linguistic interactions with generalizable techNiques—A Python library." *Psychological Methods*, vol. 24, no. 4, p. 419–438, 2019.

[41] A. Ward and D. Litman, "Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora," in *Speech and Language Technology in Education (SLaTE 2007)*, 2007, pp. 57–60.