Bilingualism: Language and Cognition

cambridge.org/bil

Producing code-switches: Adaptation of cognitive control in code-switching

Demi Zhang[1], Emiliana Pulido[1], Maria Josefina Estrada[1], Souad Kheder[2] and Edith Kaan[1]

[1]Linguistics, University of Florida, USA and [2]University of Algiers 2, Algeria

Research Article

Cite this article: Zhang, D., Pulido, E., Estrada, M.J., Kheder, S. and Kaan, E. (2026). Producing code-switches: Adaptation of cognitive control in code-switching. *Bilingualism: Language and Cognition* 1–11. https://doi.org/10.1017/S1366728926101953

Received: 23 February 2026 UTC
Revised: 10 September 2026 UTC
Accepted: 11 September 2026 UTC

Keywords:
code-switching; executive function; adaptive control; production

Corresponding author:
Edith Kaan;
Email: kaan@ufl.edu

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

A prevalent hypothesis in bilingual psycholinguistics is that processing code-switches engages domain-general cognitive control mechanisms, which should facilitate cross-task conflict adaptation. However, evidence for such cross-domain transfer remains mixed, and most prior work has focused on comprehension. The present study tested whether producing dense code-switches influences domain-general conflict control. Spanish–English bilinguals completed a cross-task adaptation paradigm in which they read aloud dense code-switched or unilingual Spanish sentences interleaved with Flanker trials. Whereas production differences were evident – articulation was slower in the dense code-switched than unilingual sentences – Flanker results showed a robust conflict effect without modulation by the preceding language context. These findings provide no evidence that producing dense code-switches induces inhibitory engagement that generalizes to subsequent non-linguistic conflict resolution. Instead, these results are compatible with the idea that dense code-switching involves control processes that do not readily transfer outside the linguistic domain and highlight methodological constraints on detecting cross-task adaptation in production.

Highlights

- Extends cross-task adaptation paradigms to dense code-switching during production
- Sentences with dense code-switches are produced slower than unilingual sentences
- Language context does not modulate Flanker conflict effects
- Suggests language control processes do not generalize to domain-general tasks

1. Introduction

Bilingual speakers frequently engage in code-switching (CS), the alternation between two languages within a single interaction or utterance. CS is a systematic phenomenon governed by grammatical, pragmatic and discourse constraints, rather than a sign of deficient language control (Muysken, 2000; Valdés Kroff, 2016). Because bilinguals must coordinate two co-activated language systems during CS, this behavior has been regarded as a natural testing ground for theories of bilingual language control and its potential relationship to domain-general executive functions (EF), such as conflict monitoring and inhibition.

Interest in the cognitive consequences of bilingual language use originally emerged from the broader "bilingual advantage" literature, which proposed that bilingual experience enhances EF (Bialystok et al., 2005). However, findings in this literature have been highly inconsistent, with large-scale reviews and meta-analyses reporting mixed or null effects (e.g., Gunnerud et al., 2020; Lehtonen et al., 2018). In addition, concerns have been raised about publication bias in the bilingual advantage literature, with studies reporting positive effects historically being more likely to appear in the published record than studies reporting null results (de Bruin et al., 2015). In response, recent work has shifted away from treating bilingualism as a categorical trait and toward examining how specific bilingual experiences, including patterns of language use and CS, shape moment-to-moment engagement of cognitive control (Bialystok & Craik, 2022; Semenova et al., 2026). The present study adopts this experience-based perspective and focuses on how CS during production influences subsequent cognitive control using a reading aloud paradigm.

1.1. Interactional contexts and adaptive language control

One of the most influential frameworks linking CS to cognitive control is the Adaptive Control Hypothesis (ACH; Green & Abutalebi, 2013). ACH proposes that bilinguals' language control systems adapt to the demands imposed by their interactional environments. The model distinguishes three interactional contexts: single-language, dual-language and dense CS contexts. In the present study, we focus specifically on dense intra-sentential CS, following the terminology of the ACH.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (http://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

In a single-language context, bilinguals use one language per environment (e.g., one language at home and another at work). Although both languages remain active, control demands largely involve goal maintenance and interference monitoring. In a dual-language context, both languages are used within the same environment, typically with different interlocutors. This context places heavier demands on cognitive control, including selective response inhibition, cue detection, task disengagement and task switching, as bilinguals must rapidly alternate between language task schemas.

Crucially, the ACH distinguishes these two contexts from the dense CS context, which is the focus of the current study and involves frequent intra-sentential mixing of both languages. Within the broader CS literature, two other intra-sentential patterns are commonly described: alternation, in which a switch occurs at a clause or phrase boundary (e.g., switching entire syntactic chunks), and insertional CS, in which lexical items or short phrases from one language are inserted into the morphosyntactic frame of the other (Belazi et al., 1994; Muysken, 2000). According to Green and Wei (2014), both alternation and insertional CS rely on a coupled control mode in which control temporarily shifts to the other language and subsequently returns to the matrix language. By contrast, dense CS reflects an open control mode in which both language schemas remain active, and lexical items from either language may be recruited as long as grammatical and discourse constraints are satisfied. Because dense CS minimizes repeated switching between language task schemas, it may rely less on sustained inhibitory control, even though competition between co-activated lexical items may still occur at local planning stages (Green, 2018).

ACH thus predicts that different interactional contexts engage distinct control processes and may differentially impact domain-general EF. In particular, dense CS may place lower demands on inhibition than alternation or insertional switching, potentially leading to weaker transfer to non-linguistic inhibitory tasks.

1.2. Cross-task adaptation approaches to CS and cognitive control

Most early studies examining CS and EF operationalized bilingual experience as a stable trait and used between-subject correlations between language background measures and EF task performance (e.g., Beatty-Martínez et al., 2020; Prior & Gollan, 2011). Such designs have interpretive challenges, including confounds related to group differences that are not controlled for (e.g., socioeconomic status, length of immersion, education or language dominance). Furthermore, such studies cannot establish causal relationships (Salig et al., 2021).

A growing body of work has adopted cross-task adaptation paradigms, in which linguistic trials are interleaved with non-linguistic EF trials. Unlike broad between-group comparisons, these paradigms manipulate immediate language context within participants and test whether engagement in a particular linguistic context is associated with short-term changes in subsequent cognitive-control performance. In this way, interactional context is operationalized as a transient processing state rather than a stable bilingual trait. Although such paradigms do not establish causality in the strong sense of longitudinal training or intervention studies, they provide a more direct test of short-term coupling between language processing and cognitive control.

One of the early studies using this approach is Wu and Thierry (2013), who interleaved Flanker trials with word-level language processing in single-language and dual-language modes. Although reaction times did not differ across conditions, the dual-language mode produced a reduced Flanker conflict effect and reduced P300 amplitudes, which the authors interpreted as enhanced inhibitory control. Neuroimaging evidence further suggests partial overlap between brain regions involved in language control and non-verbal inhibition (Branzi et al., 2016; de Baene et al., 2015), providing evidence for the plausibility of cross-task transfer.

Subsequent studies extended this approach to sentence-level processing. Adler et al. (2020) interleaved Flanker trials with self-paced reading of sentences that either contained a single intra-sentential switch or were unilingual. Participants responded faster to incongruent Flanker trials following code-switched sentences than following unilingual sentences, yielding a reduced conflict effect. Although the switches in this study were intra-sentential, they were limited to a single switching site and are best characterized as alternational CS. Under ACH, this switching pattern is assumed to involve coupled control and greater reliance on inhibition, consistent with the observed facilitation in conflict resolution. Similarly, Treffers-Daller (2009) proposed a control continuum in which alternational switching recruits the greatest degree of cognitive control, insertional switching occupies an intermediate position and dense CS recruits relatively little control.

Other studies complicate this picture. Hofweber et al. (2020) incorporated language dominance into their predictions, arguing that inhibitory demands depend on which language must be suppressed. In L1-dominant German–English bilinguals, they observed the strongest Flanker conflict effects in the L2 single-language mode, consistent with the need to inhibit the dominant L1. Although they found no significant overall differences between alternation, insertion and dense CS modes in either inhibitory control or monitoring, the predictors of performance differed across these conditions. In particular, their results suggested that dense CS draws on relatively proactive control processes, whereas alternation draws on relatively reactive control, with bilinguals' dominance profiles also playing an important role in explaining EF performance. These mixed results echo broader findings summarized in recent reviews, which emphasize substantial heterogeneity in both methods and outcomes across cross-task adaptation studies (see Semenova et al., 2026, for a review).

Kheder et al. (2026) examined different types of intra-sentential CS within a cross-task adaptation paradigm during comprehension. Across four experiments, Spanish–English bilinguals processed unilingual, insertional and dense code-switched sentences interleaved with a Flanker task, with either auditory or visual presentation and in either blocked or mixed designs. Contrary to predictions derived from production-based control models (Green, 2018; Green & Wei, 2014), the authors found no evidence that processing code-switched sentences reduced subsequent Flanker conflict effects. If anything, conflict effects were larger following code-switched sentences than following unilingual sentences, and no reliable differences emerged between dense and insertional switching in most experiments.

These findings suggest that, in comprehension, processing intra-sentential CS does not reliably boost domain-general inhibitory control or monitoring in a way that transfers to a following non-linguistic task. Instead, Kheder et al. (2026) propose that code-switched sentences presented without a broader discourse context may impose additional processing demands or require attentional refocusing, leaving fewer resources available for the subsequent Flanker trial. This work also raises the possibility that cross-task adaptation paradigms may be limited in their sensitivity to language control processes during comprehension, even if CS itself involves conflict or coordination at the linguistic level.

1.3. Bilingual language control in production versus comprehension

Most cross-task adaptation studies investigating sentential CS to date have focused on comprehension. However, the Adaptive Control Hypothesis (Green & Abutalebi, 2013) proposes that EF is adaptively tuned through the demands of bilingual language use, particularly during language production, where speakers must actively select, plan and control competing linguistic representations. Although both comprehension and production in bilinguals involve co-activation of both languages (Nozari & Novick, 2017; Pickering & Garrod, 2013), evidence suggests they rely on partially dissociable control mechanisms. Neuroimaging work indicates that language switching during production recruits dorsolateral prefrontal regions associated with executive control, whereas comprehension engages anterior cingulate cortex without overlapping category-switching activations (Blanco-Elorrieta & Pylkkänen, 2016). The difference also manifests in language switch cost. Using a lexical decision task, Reynolds et al. (2016) reported asymmetric switch costs for unbalanced bilinguals in speech production, where the switch cost for L1 is larger than the switch cost for L2. By contrast, symmetric switch costs are reported in language comprehension tasks, where the cost of switching is the same for L1 and L2. Declerck et al. (2019) observed neither comprehension-based switch costs nor robust mixing costs, suggesting that reactive (switch-triggered) and proactive (context-wide) control demands during comprehension may be weaker or qualitatively different from those posited for production.

Theoretical models reflect this distinction. The ACH focuses on task-schema coordination and sustained control demands during language production, whereas the Bilingual Interactive Activation Plus (BIA+) model accounts for comprehension through bottom-up activation of orthographic and phonological representations with limited reliance on inhibition (Dijkstra & van Heuven, 2002). Li et al. (2024) directly compared silent reading and reading aloud of paragraphs containing intra-sentential switches and found that switching elicited distinct control signatures across modalities: silent reading showed early ERP switch effects but no strong dominance asymmetry, whereas reading aloud elicited more cross-language intrusion errors when switching into the dominant language and more errors on function than content words. These results imply that inhibiting the dominant language is more critical in production than in comprehension and that production recruits additional mechanisms related to lexical selection and morphosyntactic processing. Together, these findings highlight that bilingual language control during comprehension and production is not uniform, reinforcing the need to investigate CS in production-specific paradigms rather than assuming modality-independent transfer.

Despite this, only a small number of studies have examined CS and EF using production-based cross-task adaptation paradigms. Jiao et al. (2020) interleaved picture naming with a Flanker task and observed ERP evidence for enhanced inhibitory control under mixed-language conditions, though behavioral effects were absent. The authors attributed these findings to participants' limited CS experience, suggesting that mixed-language production may have been particularly effortful for this population. More naturalistic work examining spontaneous speech has found correlations between intersentential switching frequency and inhibitory control (Han et al., 2025), yet such correlational designs do not manipulate linguistic context nor probe changes in EF state in response to moment-to-moment language use.

1.4. Reading aloud as a controlled approximation to production

Most prior work on bilingual language control in production has used cued picture naming or digit naming in language-switching paradigms (e.g., Costa & Santesteban, 2004; Declerck et al., 2012; Meuter & Allport, 1999). These tasks instantiate a strongly top-down production process: speakers start from a conceptual representation (a picture, a number) and retrieve lexical items in the cued language. Switch costs – especially larger costs for switching from L2 back to L1 in unbalanced bilinguals – have been interpreted as evidence for inhibitory control, with the more dominant language requiring stronger inhibition and therefore longer reactivation (Gade et al., 2021; Green, 1998).

However, bilingual speech production is not limited to naming pictures. In many everyday situations, speakers read written materials aloud – scripts, messages or texts that may themselves contain code-switches. Reading aloud is often characterized as a more bottom-up production modality: visual word forms activate orthographic, phonological and semantic representations (Coltheart, 2006; Dijkstra & van Heuven, 2002), which are then mapped onto articulatory plans. Empirical evidence supports that reading aloud in a bilingual context engages core language control processes. Bilinguals produce language intrusion errors primarily in mixed-language reading contexts, particularly when attempting to produce dominant-language targets, indicating inhibition and monitoring of competing languages at both lexical and sublexical levels (Gollan et al., 2014; Gollan & Goldrick, 2018). These intrusion patterns, including reversed dominance effects and sensitivity to syntactic context, suggest that reading aloud recruits multiple control mechanisms, including inhibition, default-language selection and output monitoring. Converging evidence further shows robust cross-language competition and switch costs during reading aloud (Macizo et al., 2012; Slevc et al., 2016).

At the process level, reading aloud and "free" production share core components. ERP work indicates that picture naming and word reading converge on similar phonological planning and motor stages from roughly 250 ms onward, with earlier visual stages differing mainly in intensity rather than kind (Valente et al., 2016). Clinical evidence from phonological output lexicon anomia likewise suggests a tight link between lexical retrieval for speaking and reading aloud (Gvion & Friedmann, 2016). Both production modalities are incremental: speakers typically plan and articulate in chunks rather than full sentences, though the degree of incrementality may differ. Ganushchak and Chen (2016) showed that planning during sentence reading is in some respects even more incremental than during picture description, likely reflecting differences in conceptual preparation demands.

On the other hand, Lee et al. (2026) demonstrated that production modality shapes language control. In picture naming, which engages top-down lexical selection, switch-cost asymmetries varied with context: L1-predominant contexts induced larger L1 costs, whereas balanced contexts produced symmetrical switching. In reading aloud, where lexical and phonological activation is primarily bottom-up, switch costs were symmetrical across contexts. These findings suggest that inhibitory control emerges most strongly when production demands require conceptual preparation and lexical retrieval, and they highlight that production modality modulates when and how language context engages bilingual control.

From a methodological standpoint, implementing truly naturalistic, dense CS within a cross-task adaptation paradigm is difficult, because the paradigm requires tight timing and trial-by-trial

control over linguistic input. Adopting a reading aloud paradigm therefore provides a principled compromise: it engages core production and control mechanisms that are comparable to free speech, while allowing us to precisely manipulate dense CS versus unilingual contexts and to measure their immediate carry-over effects on a subsequent Flanker task.

1.5. The current study

Building on recent comprehension-based cross-task adaptation studies by Kheder et al. (2026), which found no evidence that processing dense or insertional CS facilitates domain-general cognitive control, the present study extends this line of inquiry to the domain of language production, as influential models of bilingual language control, including the Adaptive Control Hypothesis and related control-process accounts (Green & Abutalebi, 2013), are formulated primarily for production. Specifically, we extend the dense CS condition used in Kheder et al. (2026) by examining whether producing such switches influences subsequent Flanker performance.

The study employs a reading aloud paradigm in which participants produce either dense code-switched or unilingual sentences that are interleaved trial-by-trial with a Flanker task. While reading aloud, participants engage core production mechanisms such as phonological encoding and articulatory planning, while the experimenter can maintain precise control over linguistic input and timing, constraints that are essential for cross-task adaptation paradigms. Although sentences are presented in isolation and without extended discourse context, this design enables a direct test of whether production-related language control mechanisms can modulate subsequent non-linguistic executive control under tightly controlled conditions.

We evaluate two competing hypotheses. Hypothesis A, derived from control-process accounts of dense CS (Green & Wei, 2014), predicts that dense CS involves relatively low inhibitory demands due to sustained joint activation of both languages, resulting in increased interference and a larger Flanker conflict effect compared to unilingual production. An alternative view is that, even though reading aloud engages key components of speech production, it does not constitute spontaneous speech and therefore may not fully engage the open control mode typically proposed for dense CS in natural production. Instead, reading aloud requires bilingual speakers to coordinate two languages under externally provided lexical input and to prevent intrusions during articulation, processes associated with increased language control. Under this hypothesis (B), producing dense CS may elicit a more controlled processing state, predicting a smaller Flanker conflict effect following a code-switched sentence than following a unilingual sentence (Adler et al., 2020). Importantly, both hypotheses assume that control processes engaged during language production may carry over to subsequent Flanker performance within a cross-task adaptation paradigm. However, they differ in the extent to which dense CS is expected to recruit such control processes. By contrasting dense CS and unilingual production within an interleaved sentence–Flanker paradigm, the present study tests whether language production induces control states that transfer to domain-general EF.

2. Methods

This study was preregistered at https://osf.io/q9sx3/. Materials and the dataset containing all primary study variables can be found on OSF (https://osf.io/8n7eu).

2.1. Participants

A total of 43 Spanish–English bilingual individuals were recruited for this study, primarily from the University of Florida (USA). Eligibility required participants to have acquired Spanish before English or both languages simultaneously, learned both languages before the age of 12, to report no hearing or reading impairments and to be between 18 and 35 years old. Seven participants were excluded from the analysis in accordance with our preregistered inclusion and exclusion criteria: two were excluded for scoring below 80% on comprehension questions, four were excluded because backtracking occurred on more than 40% of experimental trials,[1] one was excluded because they reported learning English at age 14, and one was excluded because they reported learning English before Spanish (who also had more than 60% backtracks). The final sample therefore consists of 36 participants.[2]

The language backgrounds of these 36 participants are summarized in Table 1, based on responses to the Bilingual Language Profile

Table 1. Language background information of the 36 participants included in the analysis

	English	Spanish
AoA	3.85 (0–12)	0
Self-reported proficiency (6 being full score)		
Listening	5.96 (5–6)	5.74 (4–6)
Reading	5.93 (5–6)	5.30 (4–6)
Writing	5.78 (4–6)	4.85 (1–6)
Speaking	5.70 (5–6)	5.26 (2–6)
Collapsed mean	5.84 (5.25–6)	5.29 (3.25–6)
Variety of spoken Spanish	Mixed varieties: 27.8%; Colombian: 16.7%; Cuban: 16.7%; Venezuelan: 16.7%; Mexican: 8.3%; Argentinian: 5.6%; Costa Rican: 2.8%; Nicaraguan: 2.8%; Peruvian: 2.8%.	

Note: values are reported as means, with ranges shown in parentheses.

[1]This participant-level exclusion criterion was not preregistered. Backtracking was preregistered as a trial-level exclusion. After inspecting data quality, we excluded participants with backtracking on more than 40% of experimental trials to remove cases with pervasive disruption during reading aloud. This threshold was selected to identify extreme cases.

[2]The preregistered sample size was established via simulation-based power analysis using data from Kheder et al. (2026, Experiment 3). The simulations indicated that 30 participants are enough to provide 80% power to detect the Flanker congruency × switching interaction. Because some participant exclusions were anticipated based on our preregistered criteria, we planned to recruit beyond this minimum target to ensure that the final analyzable sample would meet or exceed 32. Data collection was therefore stopped at 43 recruited participants, which resulted in 36 participants remaining after exclusions. As these 36 participants all met the final inclusion criteria, we retained all eligible datasets in the analyses rather than excluding additional participants to match the preregistered minimum exactly. The additional 4 datasets did not affect our results and interpretation, as the effects of interest were not significant with 36 datasets (and would not have been significant with 32). Additional details, including simulation output, are available on OSF at https://osf.io/8n7eu/files/rg7b2.

(BLP; Birdsong et al., 2012) and the Bilingual Code-Switching Profile (BCSP; Olson, 2024) questionnaires, which participants completed after the main study. These measures were selected to facilitate comparisons with samples from prior research. The BLP score, ranging from −218 (Spanish dominant) to 218 (English dominant), averaged 14.67 (*SD* = 48.86, min = −73.65, max = 130.77), indicating large variance in participants' language dominance. Meanwhile, the BCSP score, which assesses CS habits, exposure and attitudes on a 0–100 scale, averaged 66.81 (*SD* = 11.68, min = 38.47, max = 87.08), reflecting some familiarity with CS.

Before the main experiment, participants completed a bilingual verbal fluency task (Benton et al., 1994) to evaluate their ability to retrieve and produce words efficiently in each of their languages under time constraints. Participants named as many items as possible within four semantic categories per language, with 30 seconds allotted per category, in both English and Spanish. Category–language pairings were counterbalanced across participants to equate task demands across languages. Scores reflect the total number of unique items produced across the four categories within each language. On average, participants produced 43.1 items in Spanish (*SD* = 7.06, min = 28, max = 57) and 49.3 items in English (*SD* = 8.94, min = 31, max = 64).

2.2. Materials

The materials were adapted from Kheder et al. (2026), Experiment 3 and included only the unilingual Spanish and dense CS sentence conditions. An example is shown in Table 2. Crossing sentence type with Flanker congruency yielded four experimental conditions: unilingual Spanish sentences with congruent Flankers, unilingual Spanish sentences with incongruent Flankers, dense CS sentences with congruent Flankers and dense CS sentences with incongruent Flankers. Each condition contained 12 critical sentence-to-Flanker pairs, yielding 48 critical experimental pairs in total.

All sentences began in Spanish and were either entirely unilingual or contained multiple switches into English (dense CS). Individual switched segments varied in constituent size, ranging from single words to multiword constituents and occurred across a range of syntactic categories, including verbs, nouns, conjunctions, prepositions, determiners and auxiliaries. These switch sites represent categories attested in Spanish–English CS (Belazi et al., 1994; Di Sciullo et al., 1986). Thus, the present study was designed to examine the consequences of producing a dense code-switch sentence as a whole, rather than the local processing effects of particular switch types or syntactic locations. Consistent with Kheder et al. (2026), English-only sentence contexts were not included in order to reduce study duration and to prevent participants from anticipating the presence or absence of CS based on sentence-initial language. Sentence length averaged 15 words (range: 10–22). Norming studies assessing sentence plausibility and the naturalness of CS are reported in Kheder et al. (2026). Experimental items were distributed across four participant lists using a Latin-square design, such that each participant saw all four conditions but only one version of each item. The full list of experimental items is available on OSF at https://osf.io/8n7eu/files/tge4j.

In addition to the experimental items, 80 fillers (same for each list) were included to reduce predictability and maintain task engagement. These comprised 32 filler sentences (24 unilingual Spanish and 8 dense code-switched), each followed by a comprehension question and 48 filler Flanker trials. Participants therefore saw 128 total trials (48 critical sentence-to-Flanker pairs, 32 filler sentences and 48 filler Flanker trials). Participants were randomly assigned to one of the four lists. Item order was pseudorandomized, such that there were no two consecutive critical dense trials.

Table 2. Examples of the unilingual and dense code-switching conditions

Unilingual Spanish	El gato estuvo tirado en el suelo tomando sol y descansando después de la cena de la tarde.
Dense code-switching	El gato estuvo tirado *on the floor* tomando sol *and resting* después de la cena de la tarde.

Note: The cat was lying on the floor sunbathing and resting after the afternoon meal.

2.3. Procedure

Participants first provided informed consent before entering a sound-attenuated recording booth. The session began with a bilingual verbal fluency task, followed by the main experimental task.

The study was constructed and data were collected using E-Prime 3 (Psychology Software Tools, Inc.). Trials in the main task were structured as follows. Participants initiated each trial by pressing a button on a Chronos response box (Psychology Software Tools, Inc.). Each trial began with a 500 ms fixation cross presented at the center of the screen. After that, a sentence was displayed on a computer screen in its entirety. Sentences were presented in black text on a light gray background and appeared on a single line. Participants read each sentence aloud at their natural pace while their speech was recorded. Once they finished reading, they indicated completion by immediately pressing the center button on the Chronos response box.

After pressing the Chronos center button at the sentence offset, participants encountered either a Flanker trial, a comprehension question or another sentence. However, a critical sentence was always immediately followed by a critical Flanker trial. Flanker trials consisted of a central target arrow flanked by arrows pointing either in the same direction (congruent) or in the opposite direction (incongruent). Flanker trials began with a 500 ms fixation cross, followed by a five-arrow stimulus displayed at the center of the screen. The string of five arrows measured 1.5 cm in width. The Flanker stimulus remained visible until participants responded by indicating the direction of the central arrow using the response box.

Comprehension questions were preceded by a 500 ms fixation cross and remained on the screen until a response was made. Participants responded true or false by pressing either the first (Yes) or the fifth (No) button on the Chronos response box. Figure 1 illustrates the structure of a representative trial. After completing the main task, participants filled out a debriefing form and completed the BLP and BCSP language background and CS questionnaires.

2.4. Preprocessing

All recorded utterances were annotated by two Spanish–English bilingual annotators with respect to the preregistered exclusion criteria. Annotations focused on overall audibility, the presence and appropriateness of CS, and the occurrence of pauses or self-corrections.

Trials were excluded based on the preregistered criteria. Trials with reaction times on the Flankers that were shorter than 100 ms or longer than 2000 ms accounted for a small proportion of exclusions (8 trials; 0.5%).[3] Trials with incorrect responses on the

[3]As preregistered, no additional participant-wise SD-based trimming was applied. As a descriptive sensitivity check, applying a participant-wise mean ± 3-SD criterion to reciprocal RTs would have excluded 9 trials.

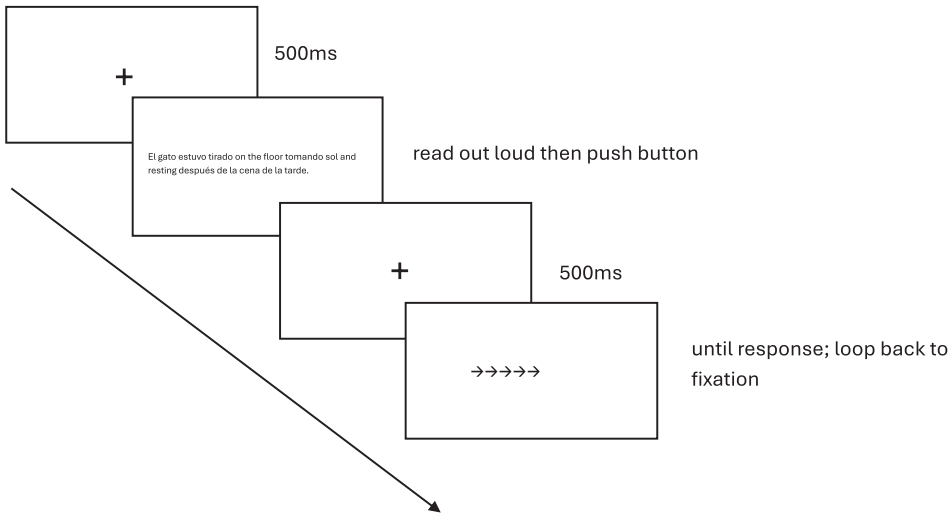


Figure 1. XXX.

Flanker task were also excluded (2 trials; 0.1%). A further set of trials was excluded because the produced utterance did not match the intended sentence condition, including cases where CS occurred in unilingual Spanish sentences or failed to occur in dense CS sentences (14 trials; 0.8%). Trials in which only a single switch was produced in the dense CS condition were excluded as well (45 trials; 2.6%). The largest proportion of excluded trials involved self-corrections following an error (also referred to as backtracks), which led to the exclusion of 275 trials (16.0%). Finally, trials were excluded when the utterance was insufficiently clear to allow reliable coding (16 trials; 0.9%). The dataset initially comprised 1,712 trials. After applying all exclusion criteria, 1,368 trials (79.9%) were retained for analysis. The number of remaining trials was very similar across the conditions (Congruent-Dense: 338; Congruent-Unilingual: 348; Incongruent-Dense: 342; Incongruent-Unilingual: 340), with each condition representing ~25% of the data.

3. Analysis and results

3.1. Flanker RT

Reaction times (RTs) were analyzed to assess the effects of Flanker condition (congruent versus incongruent) and linguistic context (dense CS versus unilingual) on Flanker performance. Mean RTs for each condition are shown in Figure 2. Participants responded more slowly in the incongruent condition ($M = 747$ ms, $SD = 251$ for Dense

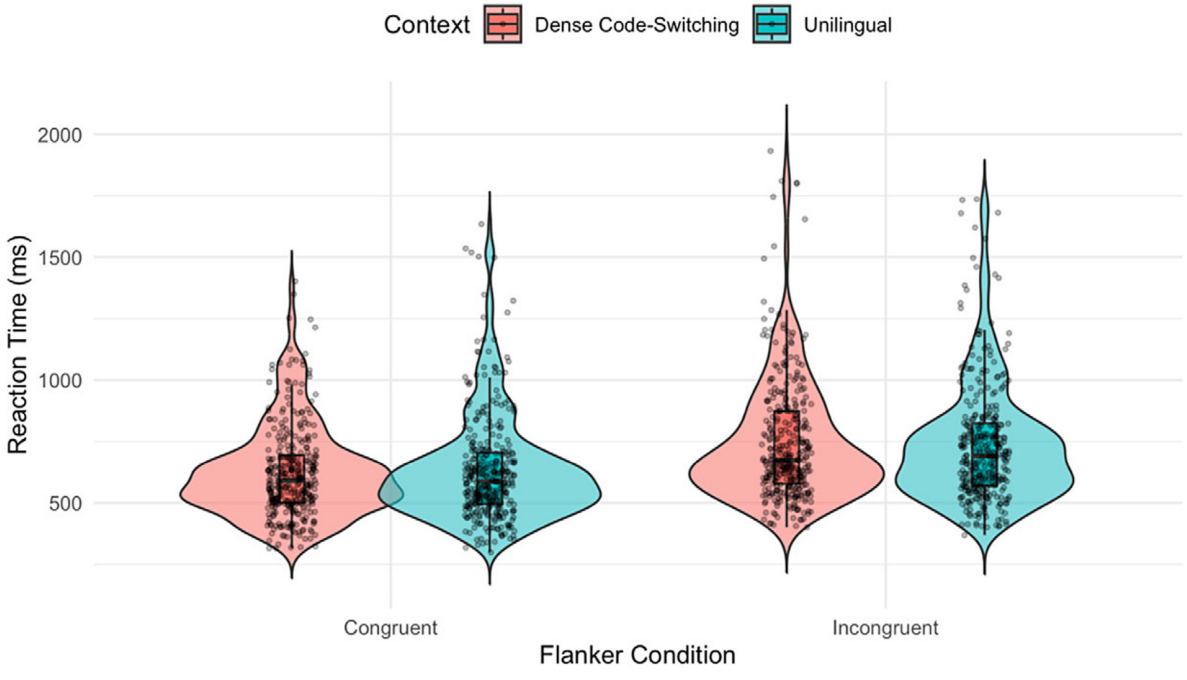


Figure 2. XXX.

CS; M = 737 ms, SD = 245 for Unilingual) than in the congruent condition in either language context (M = 624 ms, SD = 189 for Dense CS; M = 635 ms, SD = 218 for Unilingual).

Reaction time distributions were right-skewed, with a small number of very slow responses. A Box–Cox analysis (Box & Cox, 1964) was therefore used to identify an appropriate transformation. The estimated parameter ($\lambda \approx -0.7$) indicated that a reciprocal-type transformation was most suitable, performing better than log or square-root alternatives. Reaction times were thus transformed using −1000/RT. The negative sign preserves the intuitive direction of effects (with faster responses corresponding to larger transformed values), and the scaling factor of 1000 yielded coefficients of a more interpretable magnitude.

A linear mixed-effects model was then fitted using the lmer function from the lme4 package (Bates et al., 2015) running in R Version 4.4.2 (R Core Team, 2024). The model had as fixed effects: Flanker condition, Context and their interaction. To reduce collinearity and facilitate interpretation of main effects, both predictors were deviation-coded with values of −0.5 and + 0.5, with congruent Flanker and dense CS condition being coded as negative.

Initially, the model included a maximal random-effects structure with by-participant random slopes for both main effects and their interaction.[4] However, the model failed to converge and yielded near-zero variance estimates for some random slopes. To achieve a stable fit, random effects with negligible contribution were sequentially removed until convergence was obtained, resulting in a final model with a random intercept for participant.

All p-values were obtained using the Satterthwaite approximation from the lmerTest package (Kuznetsova et al., 2017). The final model revealed a significant main effect of Flanker condition, $B = 0.26$, $SE = 0.017$, $t = 15.66$, $p < .001$, confirming that incongruent trials elicited slower responses than congruent ones. The main effect of Context was not significant (B = −0.013, SE = 0.017, t = −0.77, p = .44), nor was the Flanker × Context interaction ($B = -0.017$, $SE = 0.033$, $t = -0.50$, $p = .63$). The complete outcomes of the mixed-effects models can be found in the additional online materials at https://osf.io/8n7eu/.

Together, these results demonstrate a reliable Flanker effect, indicating increased processing cost in incongruent trials, but no evidence that this effect was modulated by linguistic context. That is, the degree of CS in the speech context did not significantly influence participants' performance on the Flanker task.

3.2. Flanker performance over time

Following Adler et al. (2020), we conducted an exploratory analysis including session Half (first versus second half of the experiment) as a fixed effect to test whether conflict adaptation changed over time, under the assumption that monitoring demands might increase as participants encountered more CS throughout the session.

The linear mixed-effects model included Flanker condition, Context and session Half as fixed effects, as well as a random intercept for subject. By default, session Half was treatment-coded, with "First" coded as 0 (reference level) and "Second" coded as 1. The analysis revealed a significant main effect of session Half, $B = -0.12$, $SE = 0.02$, $t = -7.60$, $p < .001$, indicating that participants responded faster overall in the second half of the experiment (see Figure 3), consistent with a within-task learning or practice effect. The two-way interaction between Flanker condition and session Half ($B = 0.04$, $SE = 0.03$, $t = 1.33$, $p = .19$), or between Context and session Half (B = −0.02, SE = 0.03, t = −0.56, p = .58), was not significant. Likewise, the three-way interaction between Flanker condition, Context and session Half was not significant, $B = -0.05$, $SE = 0.07$, $t = -0.82$, $p = .41$. Thus, although participants became faster over time, there was no evidence that conflict adaptation either increased or decreased across the session, even as exposure to CS accumulated.

3.3. Effects of proficiency and code-switching

Given prior evidence linking bilingual proficiency and switching experience to cognitive control (Hofweber et al., 2020), we explored whether individual differences in language dominance and CS frequency modulated performance in our task. These analyses are exploratory in nature, as the present study is not sufficiently powered to test individual-difference effects or interactions involving these variables. No multiple comparison correction was used.

Language dominance (BLP) correlated positively with English fluency measures (r = .33, p < .05) and negatively with Spanish fluency (r = −.56, p < .01). Recall that a negative BLP indicates Spanish dominance and a positive indicates English dominance, so this correlation is in the predicated direction. Fluency tasks measure performance-based lexical retrieval, while BLP reflects metacognitive judgments shaped by confidence and context. Thus, the subjective and objective aspects are shown to be correlated. BLP and BCSP were not correlated (r = .02, p = .90).

Centered BLP and BCSP scores were added as continuous predictors to the main mixed-effects model, along with their interactions with Flanker condition and Context. The model included a random intercept for participant. Neither BLP nor BCSP showed significant main effects on response times (BLP: B = −0.02, SE = 0.06, p = .69; BCSP: B = −0.02, SE = 0.06, p = .79), nor did they interact with Flanker condition or Context (all p > .30).

3.4. Production quality and production duration across contexts

We also conducted exploratory analyses of production behavior, that is, error counts and backtracks, to determine whether dense CS affected how participants produced the experimental sentences.[5] Reading aloud sentences with code-switches could involve a higher processing cost, potentially resulting in increased errors, backtracks or longer durations relative to unilingual production. This is compatible with our Hypothesis B, which states that producing dense code-switched sentences in a reading aloud task recruits enhanced language control. Production errors and backtracks were manually coded on each trial and analyzed using generalized linear mixed-effects models. In the negative binomial model, the number of production errors did not reliably differ between dense CS and unilingual contexts, $B = 0.14$, $SE = 0.07$, $z = 1.90$, $p = .06$. Similarly,

[4]Following our preregistration and the analytic approach in Kheder et al. (2026), our initial model does not include a random intercept for items, as the Flanker stimuli are not expected to systematically vary in difficulty across item identities. A model including a random intercept for sentence item showed negligible item-level variance ($\tau_{00} = 0.003$) compared with subject-level variance ($\tau_{00} = 0.113$) and did not change the pattern of fixed effects (see OSF for the model and results).

[5]We define total error counts as any deviation from the original sentence, which include backtracks. However, we did separate analysis for backtracks because of its unique relationship to difficulties in resolving cross-language competition, compared with other error forms such as pauses (Korko & Williams, 2017).

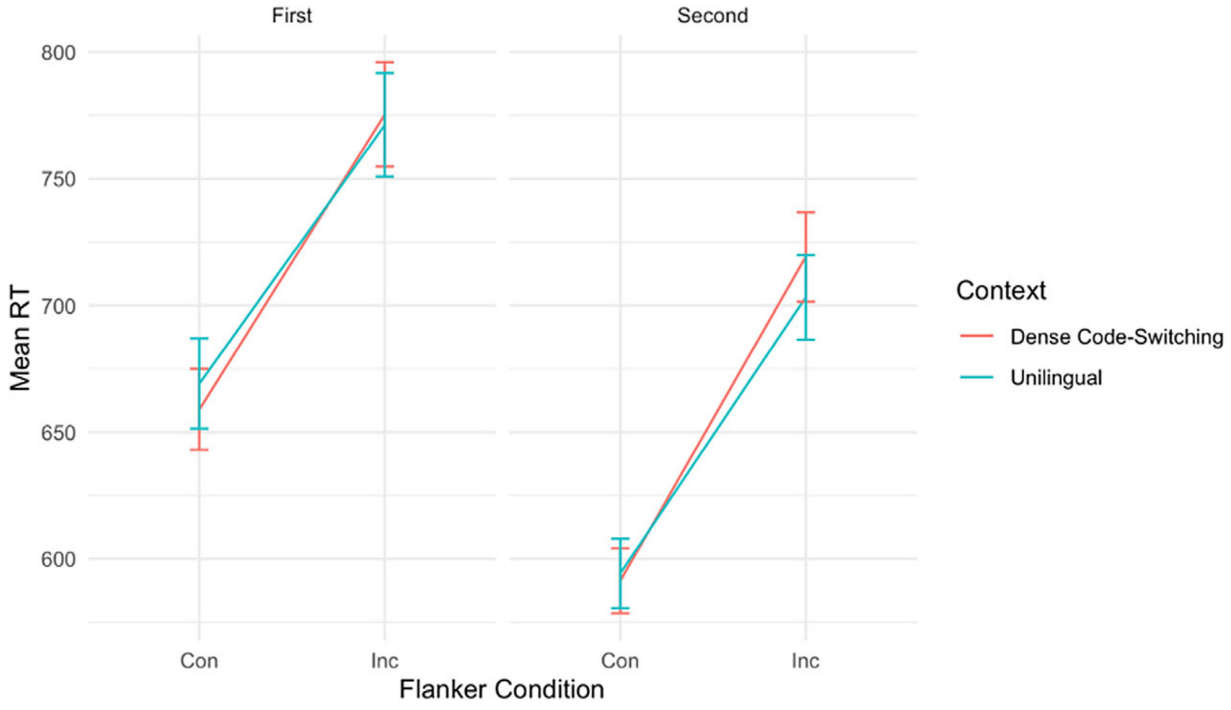


Figure 3. XXX.

using a Poisson model, the number of backtracks did not significantly differ between contexts, $B = 0.20$, $SE = 0.11$, $z = 1.86$, $p = .06$.[6] These results provide no robust evidence that reading aloud dense CS introduces additional overt production disfluencies beyond those observed in unilingual production.

In addition, we explored whether reading rate was different between CS sentences and unilingual sentences. To account for differences in sentence length, speech rate was quantified as milliseconds per syllable. Syllable counts for each experimental sentence were obtained by a fluent Spanish–English bilingual annotator through manual counting. Sentence durations were derived using the Montreal Forced Aligner (McAuliffe et al., 2017), which aligned each spoken utterance with its corresponding target sentence. Silence preceding and following the articulated sentence was excluded, yielding the duration of spoken material only. Speech rate was then computed by dividing sentence duration (in milliseconds) by the corresponding syllable count. Speech rate was analyzed using a linear mixed-effects model predicting milliseconds per syllable from context (CS or unilingual), with random intercepts and random slopes for context by participant and item. The model revealed a significant main effect of context. Unilingual Spanish sentences were articulated more quickly (M = 180.09 ms/syllable) than sentences containing code-switches (M = 202.69 ms/syllable), showing a 21.47 ms per syllable reduction ($B = -21.47$, $SE = 3.48$, $t = -6.17$, $p < .001$) in articulation time from dense CS to unilingual production. This pattern suggests that language context influences articulation duration, with code-switched sentences imposing greater production demands.

4. Discussion

The current study investigated whether producing dense code-switched sentences engages cognitive control differently from producing unilingual sentences, and whether any such differences carry over to performance on a subsequent non-linguistic (domain-general) conflict task. Production-based models of CS have proposed that different language contexts place distinctive demands on language control mechanisms (Green & Abutalebi, 2013; Green & Wei, 2014). Specifically, it has been argued that dense CS relies on a sustained co-activation of both languages, reducing the reliance on repeated shifts between re-engaging and disengaging language-level control configurations (Green & Wei, 2014). Although most prior studies have focused on comprehension rather than production, bilingual production and comprehension involve different control processes. To our knowledge, the current study is the first to test the impact of CS production on cognitive control using a cross-task adaptation paradigm. We contrasted two predictions: either dense CS weakens transient control (predicting a larger conflict effect) or it engages sustained control (predicting a smaller conflict effect).

Contrary to both of our hypotheses, we found no evidence that language context modulated performance on the Flanker task. There was no language context effect: response times did not differ following dense code-switched compared to unilingual production, and language context did not interact with Flanker congruency. However, we did observe the expected congruency effect, with slower response times on incongruent than congruent trials independent of language context. In addition, dense code-switched

[6]Because error counts and backtracks are non-negative integers and exhibit a skewed distribution with many zeros, we first fit a Poisson model with random intercepts for subjects and items. This model showed substantial overdispersion (Pearson $\chi^2 = 1975.67$, ratio = 1.16, $p < .001$), and therefore a negative binomial model was adopted for error counts. On the other hand, no overdispersion was confirmed for backtracks (Pearson $\chi 2 = 1452.33$, ratio = 0.85, $p = .99$), therefore the Poisson model was used.

sentences were articulated more slowly than unilingual sentences after controlling for linguistic length (ms/syllable), suggesting increased planning and coordination demands during bilingual production.

We acknowledge that we cannot draw strong conclusions on the basis of the null effects, but we would like to discuss two interpretations of the absence of a language context effect on the Flanker performance. First, there may not be a difference in cognitive control engagement between CS and unilingual production in our study. This is in line with production-based models of bilingual language control, according to which dense CS has a limited reliance on inhibitory regulation (Green, 2018; Green & Abutalebi, 2013).

In the present study, we observed no differences in backtracking across language contexts (unilingual Spanish versus dense code-switched sentences). Because backtracking has been linked to monitoring and repair processes (Korko & Williams, 2017), the absence of context differences suggests that dense CS was not treated as a conflict requiring additional repair or control mechanisms. At the same time, the overall slowdown observed for code-switched sentences does not necessarily imply increased domain-general cognitive control demands. Instead, it may reflect processing costs at other stages of production. One possibility is increased phonological encoding load: switching between Spanish and English phonological systems may require reconfiguration of articulatory planning, leading to longer preparation times. The slowdown may therefore arise from motor planning and articulatory execution rather than conflict monitoring.

Second, the observed null effect may reflect a limited sensitivity of the cross-task adaptation paradigm to linguistic-related processes. If control engagement is strongly context-dependent, then adjustments made during language production may not be carried over to a non-linguistic task with different demands. Learning-based accounts of control predict that such generalization is limited when tasks do not share overlapping control requirements (Nozari et al., 2016; Zhu et al., 2025). Nozari et al. (2016) argue that cognitive control adaptation often reflects learning at very specific representational levels, not a generalizable upregulation of control across different tasks. Therefore, absence of cross-task adaptation does not simply imply absence of control, but instead that the cognitive control that was possibly recruited for the linguistic task may not transfer over to the following non-linguistic task, since these may rely on different underlying demands. A recent extension of the Learning-based account further emphasizes that control learning is constrained by the specifications of the task structure and may generalize only across domains with a possible overlap (Zhu et al., 2025). In our case, the tasks and representations may have been too different to affect each other.

Our results do not replicate the reduced Flanker conflict effects reported by Adler et al. (2020), who used a similar paradigm and population but found facilitation following intrasentential code-switches. However, Adler et al. conducted a comprehension study in which sentences switched once from Spanish to English at varying grammatical locations, whereas the present study used dense CS stimuli with multiple switches in a production task. Differences in switch type may matter: alternational switching more closely resembles dual-language contexts, which have been argued to recruit greater inhibitory and switching demands, whereas dense CS may operate under an open control mode and therefore involve limited inhibition (Green & Wei, 2014). Similarly, Treffers-Daller (2009) proposes that alternational switching recruits greater cognitive control than dense CS, with alternational switching and dense CS occupying opposite ends of the control continuum. Consistent with this possibility, comprehension-based studies in Spanish–English bilinguals have also reported null or inconsistent conflict adaptation following dense CS (Kheder et al., 2026). Another potential explanation is methodological. Adler et al.'s facilitation effect was relatively small and emerged only in the second half of the session, raising the possibility that the original study may have been underpowered.

Our dataset showed substantial variation in language dominance across participants, with neither language clearly dominant at the group level; yet we failed to find effects of language background on the Flanker effects. This makes it unlikely that the null effect is driven by systematic dominance differences, and instead points to a possible limitation of the cross-task paradigm in capturing cognitive control transfer to a domain-general task.

5. Limitations and conclusions

One limitation of the present study is that production was operationalized through reading aloud of generated sentences, which differs from spontaneous speech. Even if the materials are made to resemble naturally code-switched sentences, a reading aloud task restricts participants in their word choice and may not encourage entering the open control mode proposed by the ACH. In addition, participants may have engaged in processing strategies and experienced task demands that differ from those involved in spontaneous conversational CS, which may have affected the observed adaptation effects. Spontaneous production tasks would be better in this respect but would be challenging to incorporate into a cross-task adaptation paradigm.

A second limitation is the lack of a unilingual English control. Our participants varied in the age of English acquisition and language dominance and hence may have differed in the ease of producing English. Even though the bilingual profile measures (BLP, BCSP) did not have any effect in our current study, including unilingual English sentences would have allowed for a better assessment of the differences in production speed between the CS condition (which included English) and the unilingual Spanish conditions. In addition, the Flanker congruency effects could have differed between the unilingual English and the conditions containing Spanish. At the same time, adding such a unilingual English condition would have made it necessary to include English-to-Spanish code-switches so that participants would not associate CS with Spanish-initial sentences only. This would have made the study considerably longer.

Another methodological consideration concerns the absence of a response deadline for the Flanker trials. As discussed in Kheder et al. (2026), unrestricted responding may attenuate cognitive-control effects that are more apparent under speeded conditions (see also Dunaway & Weissman, 2025), potentially allowing slower and more strategic responding that obscures transient inhibition effects (Moretti et al., 2026).

Many prior studies on CS and cognitive control employed comprehension tasks. The current study extends this line of research by testing production. In line with recent comprehension-based studies that found no or weak cross-domain adaptation effects, we obtained no evidence that the production of code-switches engages domain-general cognitive control differently from the production of unilingual sentences. Our work therefore contributes to the discussion of bilingual language control and its complexity.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728926101953.

Data availability statement. The data that support the findings of this study are openly available in https://osf.io/8n7eu.

Acknowledgments. The authors would like to thank Jorge Valdés Kroff, Hannah Treadway, and Alexandra Dabrowski.

Funding statement. This work was supported by the National Science Foundation (NSF BCS-2017251).

Competing interests. The authors declare no competing interests.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

AI contributions to research. ChatGPT 5.2 was used for language polishing and grammar checking purposes.

References

Adler, R. M., **Valdés Kroff, J. R.**, & **Novick, J. M.** (2020). Does integrating a code-switch during comprehension engage cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **46**(4), 741–759. https://doi.org/10.1037/xlm0000755

Bates, D., **Mächler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1). https://doi.org/10.18637/jss.v067.i01

Beatty-Martínez, A. L., **Navarro-Torres, C. A.**, **Dussias, P. E.**, **Bajo, M. T.**, **Guzzardo Tamargo, R. E.**, & **Kroll, J. F.** (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **46**(6), 1022–1047. https://doi.org/10.1037/xlm0000770

Belazi, H. M., **Rubin, E. J.**, & **Toribio, A. J.** (1994). Code switching and X-bar theory: The functional head constraint. *Linguistic Inquiry*, **25**(2), 221–237.

Benton, A. L., **Hamsher, K. D. S.**, & **Sivan, A. B.** (1994). *Multilingual aphasia examination* (3rd ed.) AJA Associates.

Bialystok, E., & **Craik, F. I. M.** (2022). How does bilingualism modify cognitive function? Attention to the mechanism. *Psychonomic Bulletin & Review*, **29**(4), 1246–1269. https://doi.org/10.3758/s13423-022-02057-5

Bialystok, E., **Martin, M. M.**, & **Viswanathan, M.** (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, **9**(1), 103–119. https://doi.org/10.1177/13670069050090010701

Birdsong, D., **Gertken, L. M.**, & **Amengual, M.** (2012). *Bilingual language profile: An easy-to-use instrument to assess bilingualism*. COERLL, University of Texas at Austin. https://blp.coerll.utexas.edu/

Blanco-Elorrieta, E., & **Pylkkänen, L.** (2016). Bilingual language control in perception versus action: MEG reveals comprehension control mechanisms in anterior cingulate cortex and domain-general control of production in dorsolateral prefrontal cortex. *The Journal of Neuroscience*, **36**(2), 290–301. https://doi.org/10.1523/JNEUROSCI.2597-15.2016

Box, G. E. P., & **Cox, D. R.** (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: (Methodological)*, **26**(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Branzi, F. M., **Della Rosa, P. A.**, **Canini, M.**, **Costa, A.**, & **Abutalebi, J.** (2016). Language control in bilinguals: Monitoring and response selection. *Cerebral Cortex*, **26**(6), 2367–2380. https://doi.org/10.1093/cercor/bhv052

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, **4**(1). https://doi.org/10.1080/13603110600574322

Costa, A., & **Santesteban, M.** (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, **50**(4), 491–511. https://doi.org/10.1016/j.jml.2004.02.002

De Baene, W., **Duyck, W.**, **Brass, M.**, & **Carreiras, M.** (2015). Brain circuit for cognitive control is shared by task and language switching. *Journal of Cognitive Neuroscience*, **27**(9), 1752–1765. https://doi.org/10.1162/jocn_a_00817

De Bruin, A., **Treccani, B.**, & **Della Sala, S.** (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, **26**(1), 99–107. https://doi.org/10.1177/0956797614557866

Declerck, M., **Koch, I.**, **Duñabeitia, J. A.**, **Grainger, J.**, & **Stephan, D. N.** (2019). What absent switch costs and mixing costs during bilingual language comprehension can tell us about language control. *Journal of Experimental Psychology: Human Perception and Performance*, **45**(6), 771–789. https://doi.org/10.1037/xhp0000627

Declerck, M., **Koch, I.**, & **Philipp, A. M.** (2012). Digits vs pictures: The influence of stimulus type on language switching. *Bilingualism: Language and Cognition*, **15**(4), 896–904. https://doi.org/10.1017/S1366728912000193

Di Sciullo, A.-M., **Muysken, P.**, & **Singh, R.** (1986). Government and code-mixing. *Journal of Linguistics*, **22**(1), 1–24. https://doi.org/10.1017/S0022226700010537

Dijkstra, T., & **van Heuven, W. J. B.** (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, **5**(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dunaway, M. G., & **Weissman, D. H.** (2025). Both congruent and incongruent trials drive the congruency sequence effect: Novel support for an episodic retrieval view of adaptive control in the prime–probe task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **51**(2), 171–189. https://doi.org/10.1037/xlm0001349

Gade, M., **Declerck, M.**, **Philipp, A. M.**, **Rey-Mermet, A.**, & **Koch, I.** (2021). Assessing the evidence for asymmetrical switch costs and reversed language dominance effects – A meta-analysis. *Journal of Cognition*, **4**(1), 55. https://doi.org/10.5334/joc.186

Ganushchak, L. Y., & **Chen, Y.** (2016). Incrementality in planning of speech during speaking and reading aloud: Evidence from eye-tracking. *Frontiers in Psychology*, **7**. https://doi.org/10.3389/fpsyg.2016.00033

Gollan, T. H., & **Goldrick, M.** (2018). A switch is not a switch: Syntactically-driven bilingual language control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **44**(1), 143–156. https://doi.org/10.1037/xlm0000462

Gollan, T. H., **Schotter, E. R.**, **Gomez, J.**, **Murillo, M.**, & **Rayner, K.** (2014). Multiple levels of bilingual language control: Evidence from language intrusions in reading aloud. *Psychological Science*, **25**(2), 585–595. https://doi.org/10.1177/0956797613512661

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, **1**(2), 67–81. https://doi.org/10.1017/S1366728998000133

Green, D. W. (2018). Language control and code-switching. *Languages*, **3**(2), 8. https://doi.org/10.3390/languages3020008

Green, D. W., & **Abutalebi, J.** (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, **25**(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Green, D. W., & **Wei, L.** (2014). A control process model of code-switching. *Language, Cognition and Neuroscience*, **29**(4), 499–511. https://doi.org/10.1080/23273798.2014.882515

Gunnerud, H. L., **Ten Braak, D.**, **Reikerås, E. K. L.**, **Donolato, E.**, & **Melby-Lervåg, M.** (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, **146**(12), 1059–1083. https://doi.org/10.1037/bul0000301

Gvion, A., & **Friedmann, N.** (2016). A principled relation between reading and naming in acquired and developmental anomia: Surface dyslexia following impairment in the phonological output lexicon. *Frontiers in Psychology*, **7**. https://doi.org/10.3389/fpsyg.2016.00340

Han, X., **Li, W.**, & **Filippi, R.** (2025). Modulating bilingual language production and cognitive control: How bilingual language experience matters. *Bilingualism: Language and Cognition*, **28**(1), 120–134. https://doi.org/10.1017/S1366728924000191

Hofweber, J., **Marinis, T.**, & **Treffers-Daller, J.** (2020). Experimentally induced language modes and regular code-switching habits boost bilinguals' executive performance: Evidence from a within-subject paradigm. *Frontiers in Psychology*, **11**, 542326. https://doi.org/10.3389/fpsyg.2020.542326

Jiao, L., **Grundy, J. G.**, **Liu, C.**, & **Chen, B.** (2020). Language context modulates executive control in bilinguals: Evidence from language production. *Neuropsychologia*, **142**, 107441. https://doi.org/10.1016/j.neuropsychologia.2020.107441

Kheder, S., **Mello Medina, R.**, **Valdés Kroff, J.**, & **Kaan, E.** (2026). Investigating the effects of code-switch types on cognitive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **52**(6), 934–953. https://doi.org/10.1037/xlm0001510

Korko, M., & **Williams, S. A.** (2017). Inhibitory control and the speech patterns of second language users. *British Journal of Psychology*, **108**(1), 43–72. https://doi.org/10.1111/bjop.12176

Kuznetsova, A., **Brockhoff, P. B.**, & **Christensen, R. H. B.** (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13). https://doi.org/10.18637/jss.v082.i13

Lee, Y.-W., **Rebuschat, P.**, & **Casaponsa, A.** (2026). Top-down and bottom-up bilingual speech production: The effects of language context on inhibitory control. *Bilingualism: Language and Cognition*, **29**(3), 686–698. https://doi.org/10.1017/S1366728925100357

Lehtonen, M., **Soveri, A.**, **Laine, A.**, **Järvenpää, J.**, **De Bruin, A.**, & **Antfolk, J.** (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, **144**(4), 394–425. https://doi.org/10.1037/bul0000142

Li, C., **Midgley, K. J.**, **Ferreira, V. S.**, **Holcomb, P. J.**, & **Gollan, T. H.** (2024). Different language control mechanisms in comprehension and production: Evidence from paragraph reading. *Brain and Language*, **248**, 105367. https://doi.org/10.1016/j.bandl.2023.105367

Macizo, P., **Bajo, T.**, & **Paolieri, D.** (2012). Language switching and language competition. *Second Language Research*, **28**(2), 131–149. https://doi.org/10.1177/0267658311434893

McAuliffe, M., **Socolof, M.**, **Mihuc, S.**, **Wagner, M.**, & **Sonderegger, M.** (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. *Interspeech 2017*, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

Meuter, R. F. I., & **Allport, A.** (1999). Bilingual language switching in naming: A symmetrical costs of language selection. *Journal of Memory and Language*, **40**(1), 25–40. https://doi.org/10.1006/jmla.1998.2602

Moretti, L., **Koch, I.**, **Hornjak, R.**, & **von Bastian, C. C.** (2026). Quality over quantity: Focusing on high-conflict trials to improve the reliability and validity of attentional control measures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **52**(1), 34–54. https://doi.org/10.1037/xlm0001466

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press.

Nozari, N., & **Novick, J.** (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, **26**(5), 403–410. https://doi.org/10.1177/0963721417702419

Nozari, N., **Trueswell, J. C.**, & **Thompson-Schill, S. L.** (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychonomic Bulletin & Review*, **23**(6), 1942–1953. https://doi.org/10.3758/s13423-016-1068-8

Olson, D. J. (2024). The bilingual code-switching profile (BCSP): Assessing the reliability and validity of the BCSP questionnaire. *Linguistic Approaches to Bilingualism*, **14**(3), 400–433. https://doi.org/10.1075/lab.21039.ols

Pickering, M. J., & **Garrod, S.** (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, **36**(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Prior, A., & **Gollan, T. H.** (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society*, **17**(4), 682–691. https://doi.org/10.1017/S1355617711000580

R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reynolds, M. G., **Schlöffel, S.**, & **Peressotti, F.** (2016). Asymmetric switch costs in numeral naming and number word reading: Implications for models of bilingual language production. *Frontiers in Psychology*, **6**. https://doi.org/10.3389/fpsyg.2015.02011

Salig, L. K., **Valdés Kroff, J. R.**, **Slevc, L. R.**, & **Novick, J. M.** (2021). Moving from bilingual traits to states: Understanding cognition and language processing through moment-to-moment variation. *Neurobiology of Language*, **2**(4), 487–512. https://doi.org/10.1162/nol_a_00046

Semenova, E., **Norkina, M.**, **Logvinenko, T.**, **Ryseva, K.**, **Chinn, L. K.**, **Crabb, K.**, & **Grigorenko, E. L.** (2026). The modulating role of interactional contexts in executive functioning of bilinguals: A scoping review. *Bilingualism: Language and Cognition*, **29**(2), 527–561. https://doi.org/10.1017/S1366728925100084

Slevc, L. R., **Davey, N. S.**, & **Linck, J. A.** (2016). A new look at "the hard problem" of bilingual lexical access: Evidence for language-switch costs with univalent stimuli. *Journal of Cognitive Psychology*, **28**(4), 385–395. https://doi.org/10.1080/20445911.2016.1152274

Treffers-Daller, J. (2009). Code-switching and transfer: An exploration of similarities and differences. In *The Cambridge handbook of linguistic code-switching* (pp. 58–74). Cambridge University Press. https://doi.org/10.1017/CBO9780511576331.005

Valdés Kroff, J. R. (2016). Mixed NPs in Spanish-English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In R. E. G.Tamargo, C. M. Mazak, & M. C. P.Couto (Eds.), *Spanish-English codeswitching in the Caribbean and the US* (pp. 281–300). John Benjamins. https://doi.org/10.1075/ihll.11.12val

Valente, A., **Pinet, S.**, **Alario, F.-X.**, & **Laganaro, M.** (2016). "When" does picture naming take longer than word reading? *Frontiers in Psychology*, **7**. https://doi.org/10.3389/fpsyg.2016.00031

Wu, Y. J., & **Thierry, G.** (2013). Fast modulation of executive function by language context in bilinguals. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, **33**(33), 13533–13537. https://doi.org/10.1523/JNEUROSCI.4760-12.2013

Zhu, D., **Wang, X.**, **Zhao, E.**, **Nozari, N.**, **Notebaert, W.**, & **Braem, S.** (2025). Cognitive control is task specific: Further evidence against the idea of domain-general conflict adaptation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **51**(11), 1703–1715. https://doi.org/10.1037/xlm0001480