L1 phonetic drift during adult L2 acquisition

Speech rate analysis

Joseph V. Casillas

2024-04-15

Overview

A common finding in the second language (L2) speech literature is that adults who learn another language typically retain a non-native accent in the target language (TL). L1 to L2 cross-linguistic influence is attested to some degree in learners of all languages at all proficiency levels, showing that learners tend to produce L2 sounds through the filter of their L1. With time and effort, L2 learners develop TL-specific phonetic categories. Recent research suggests that the cross-linguistic influence that occurs in bilingual speech may be bidirectional, that is to say, L2 phonetic category formation may have behavioral consequences for L1 categories. This project explores the production of native English speaking adult L2 learners of Spanish. Specifically, it tracks the phonological acquisition of stop contrasts over the course of a domestic immersion program in order to shed light on how L2 phonological acquisition affects L1 phonology.

The problem

- Learning to produce and perceive L2 sounds is hard
- During early stages of learning we produce target language sounds using L1 phonology

Theoretical issues

Opposing viewpoints:

- **Speech Learning Model** (SLM-r: Flege and Bohn (2021))

    - L1 and L2 phonetic categories share the same phonetic space
    - Bidirectional influence, i.e., learning L2 sounds can affect L1 sounds

- **Second Language Linguistic Perception Model** (L2LP: Van Leussen and Escudero (2015))

    - During early stages of learning L1 phonological system is copied
    - L2 sounds develop independently of L1 system

- As predicted by the SLM-r, there is growing evidence that **phonetic drift** (PD) occurs during L2 learning (See Chang 2012, 2013; Lang and Davidson 2019; Tobin, Nam, and Fowler 2017)

- PD refers to changes in L1 (native) categories that take place during the process of learning novel sounds in an L2 .

- The SLM predicts phonetic drift, the L2LP suggests it should not occur

Data prep

Load libraries and helper functions

```r
source(here::here("scripts", "r", "01_helpers.R"))

theme_set(
  theme_bw(base_size = 12, base_family = "Palatino") +
  theme(
    axis.title.y = element_text(size = rel(0.9), hjust = 0.95),
    axis.title.x = element_text(size = rel(0.9), hjust = 0.95),
    panel.grid.major = element_line(colour = "grey90", linewidth = 0.15),
    panel.grid.minor = element_line(colour = "grey90", linewidth = 0.15),
    strip.background = element_blank()
    )
  )
```

Load data

```r
spr_en <- read_csv(here("data", "raw", "speech_rate_en.csv"))
spr_sp <- read_csv(here("data", "raw", "speech_rate_sp.csv"))
```

Tidy data

```r
dat <- bind_rows(
  spr_en |>
    clean_names() |>
    separate(col = soundname, into = c("id", "language"), sep = "_") |>
    separate(col = id, into = c("id", "year"), sep = -4) |>
    separate(col = id, into = c("id", "phase"), sep = "[0-9](?=p)") |>
    mutate(language = "English"),

  spr_sp |>
    clean_names() |>
    separate(col = soundname, into = c("id", "phase"), sep = 6) |>
    separate(
      col = phase,
      into = c("phase", "year"),
      sep = "(?<=[a-zA-Z])\\s*(?=[0-9])"
    ) |>
    mutate(language = "Spanish") |>
    select(id, phase, year, language, everything())
) |>
  rename(
    n_syllables = nsyll,
    n_pauses = npause,
    recording_duration = dur_s,
    phonation_time = phonationtime_s,
    sr = speechrate_nsyll_dur,
    ar = articulation_rate_nsyll_phonationtime,
    asd = asd_speakingtime_nsyll
  ) |>
  pivot_longer(
    cols = n_syllables:asd,
    names_to = "metric",
    values_to = "val"
  ) |>
  mutate(
    phase = fct_relevel(phase, "pre"),
    metric = fct_relevel(metric,
    "n_syllables", "n_pauses", "recording_duration", "phonation_time",
    "sr", "ar", "asd")
  ) |>
  write_csv(here("data", "tidy", "spr_tidy.csv"))
```

```
glimpse(dat)
```

```
Rows: 840
Columns: 6
$ id       <chr> "aka04", "aka04", "aka04", "aka04", "aka04", "aka04", "aka04"~
$ phase    <fct> post, post, post, post, post, post, post, pre, pre, pre, pre,~
$ year     <chr> "2019", "2019", "2019", "2019", "2019", "2019", "2019", "2019~
$ language <chr> "English", "English", "English", "English", "English", "Engli~
$ metric   <fct> n_syllables, n_pauses, recording_duration, phonation_time, sr~
$ val      <dbl> 411.000, 36.000, 109.500, 85.370, 3.750, 4.810, 0.208, 251.00~
```

Design

Variables

- id: unique participant identification string
- phase: stage of data collection (pre, post)
- year: year of immersion program
- language: language spoken during interview (Spanish, English)

Metrics

- n_syllables: number of syllables produced
- n_pauses: number of pauses
- recording_duration: total time (s) of recording
- phonation_time: total speaking time (s)
- sr: speech rate (n_syllables / recording_duration)
- ar: articulation rate (n_syllables / phonation_time)
- asd: average syllable duration (phonation_time / n_syllables)

Exploratory plots

Spanish interview

OPI-style oral assessment in which participants are asked a range of questions.
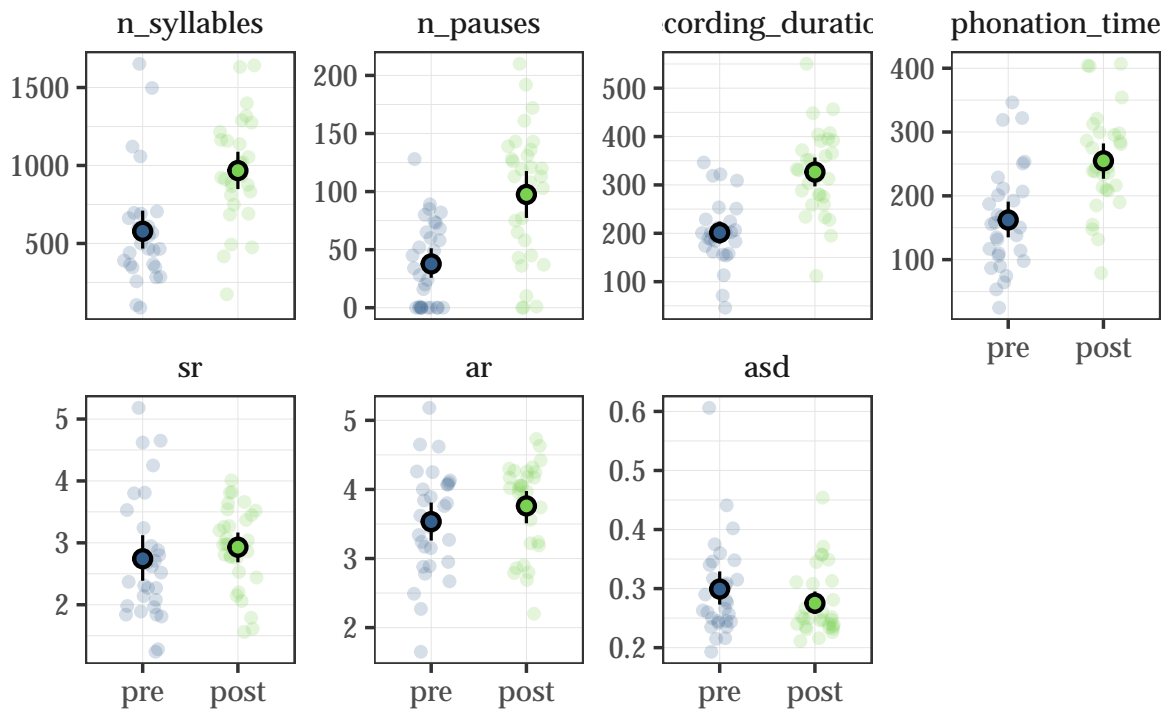
```r
dat |>
  filter(language == "Spanish") |>
  ggplot() +
  aes(x = phase, y = val) +
  facet_wrap(. ~ metric, scales = "free_y", ncol = 4) +
  geom_jitter(
    aes(color = phase),
    height = 0, width = 0.2, alpha = 0.2, show.legend = F
  ) +
  stat_summary(
    aes(fill = phase),
    fun.data = mean_cl_boot, geom = "pointrange", pch = 21,
    show.legend = F
  ) +
  scale_fill_viridis_d(option = "D", begin = 0.3, end = 0.8) +
  scale_color_viridis_d(option = "D", begin = 0.3, end = 0.8) +
  labs(
    title = "Speech fluency metrics",
    subtitle = "Spanish production as a function of phase in immersion program",
    x = NULL, y = NULL
  )
```

Speech fluency metrics

Spanish production as a function of phase in immersion program


English

Participants respond to question "what does learning another language mean to you?"

```
dat |>
  filter(language == "English") |>
  ggplot() +
  aes(x = phase, y = val) +
  facet_wrap(. ~ metric, scales = "free_y", ncol = 4) +
  geom_jitter(
    aes(color = phase),
    height = 0, width = 0.2, alpha = 0.2, show.legend = F
  ) +
  stat_summary(
    aes(fill = phase),
    fun.data = mean_cl_boot, geom = "pointrange", pch = 21,
    show.legend = F
```

```
) +
scale_fill_viridis_d(option = "D", begin = 0.3, end = 0.8) +
scale_color_viridis_d(option = "D", begin = 0.3, end = 0.8) +
labs(x = NULL, y = NULL) +
labs(
  title = "Speech fluency metrics",
  subtitle = "English production as a function of phase in immersion program",
  x = NULL, y = NULL
)
```

Speech fluency metrics
English production as a function of phase in immersion program
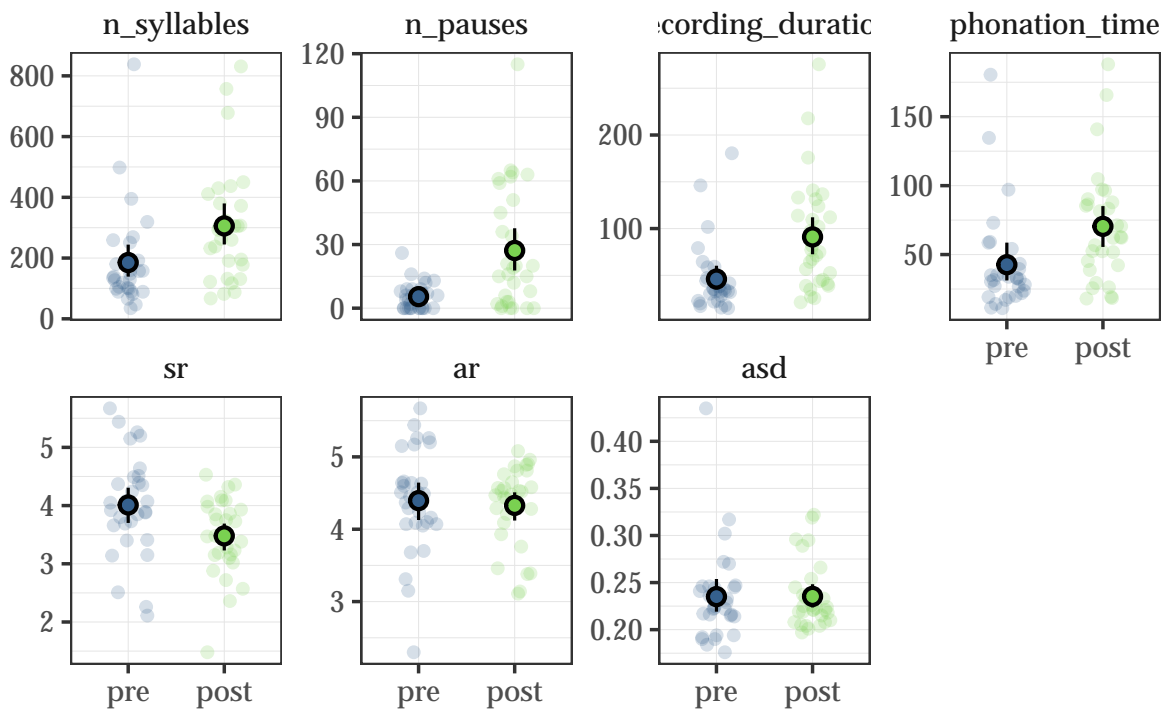
Models

```
# Scale continuous variables
dat_z <- dat |>
```

```
  group_by(language, metric) |>
  mutate(val_z = scale_this(val))

bf_raw <- bf(
  val ~ 0 + phase:language +
    (1 + phase:language | id)
)

bf_z <- bf(
  val_z ~ 0 + phase:language +
    (1 + phase:language | id)
)
```

Number of syllables

```
get_prior(
  formula = bf_raw,
  family = poisson(link = "log"),
  data = filter(dat_z, metric == "n_syllables")
) |>
  select(prior, class, coef) |>
  knitr::kable()
```

prior	class	coef
	b	
	b	phasepost:languageEnglish
	b	phasepost:languageSpanish
	b	phasepre:languageEnglish
	b	phasepre:languageSpanish
lkj(1)	cor	
	cor	
student_t(3, 0, 2.5)	sd	
	sd	
	sd	Intercept
	sd	phasepost:languageEnglish
	sd	phasepost:languageSpanish
	sd	phasepre:languageEnglish
	sd	phasepre:languageSpanish

```r
n_syllables_prior <- c(
  prior(normal(5, 1), class = "b"),
  prior(lkj(1), class = "cor"),
  prior(exponential(1), class = "sd")
)

mod_sr_n_syllables <- brm(
  formula = bf_raw,
  family = poisson(link = "log"),
  prior = n_syllables_prior,
  sample_prior = "no",
  chains = 4, iter = 2000, warmup = 1000,
  cores = 4, threads = threading(2),
  backend = "cmdstanr",
  control = list(adapt_delta = 0.99, max_treedepth = 12),
  data = filter(dat_z, metric == "n_syllables"),
  file = here("models", "mod_sr_n_syllables")
)
```

```r
n_syllables_post <- as_tibble(mod_sr_n_syllables) |>
  transmute(
    English_pre = exp(`b_phasepre:languageEnglish`),
    English_post = exp(`b_phasepost:languageEnglish`),
    Spanish_pre = exp(`b_phasepre:languageSpanish`),
    Spanish_post = exp(`b_phasepost:languageSpanish`)
  ) |>
  pivot_longer(cols = everything()) |>
  separate(name, into = c("language", "phase"), sep = "_") |>
  mutate(
    phase = str_to_title(phase),
    phase = fct_relevel(phase, "Pre")
  )

n_syllables_post |>
  ggplot() +
  aes(y = language, x = value, group = phase) +
  geom_point(
    data = filter(dat_z, metric == "n_syllables") |>
      mutate(phase = str_to_title(phase), phase = fct_relevel(phase, "Pre")),
    mapping = aes(y = language, x = val, fill = phase),
    position = ggpp::position_dodgenudge(
      x = 0, direction = "split", width = 0.5
```

```
  ),
  pch = 21, alpha = 0.4, size = 3, color = "white", stroke = 1
) +
stat_halfeye(
  aes(fill = phase),
  position = ggpp::position_dodgenudge(
    y = 0.06, direction = "split", width = 0.5),
  pch = 21, point_fill = "white", show.legend = F
  ) +
scale_fill_viridis_d(name = NULL, option = "D", begin = 0.4, end = 0.8) +
scale_color_viridis_d(name = NULL, option = "D", begin = 0.4, end = 0.8) +
labs(y = NULL, x = "Syllables\n") +
guides(fill = guide_legend(override.aes = list(alpha = 1, size = 3))) +
theme(
  legend.position = "inside", legend.justification.inside = c(1, 0),
  legend.background = element_blank(),
  axis.text.y = element_text(angle = 90, hjust = 0.5, size = 16),
  axis.text.x = element_text(size = 14),
  axis.title.x = element_text(size = 16),
  legend.text = element_text(size = 14)
)
```

```
Registered S3 methods overwritten by 'ggpp':
  method                 from
  heightDetails.titleGrob ggplot2
  widthDetails.titleGrob  ggplot2
```
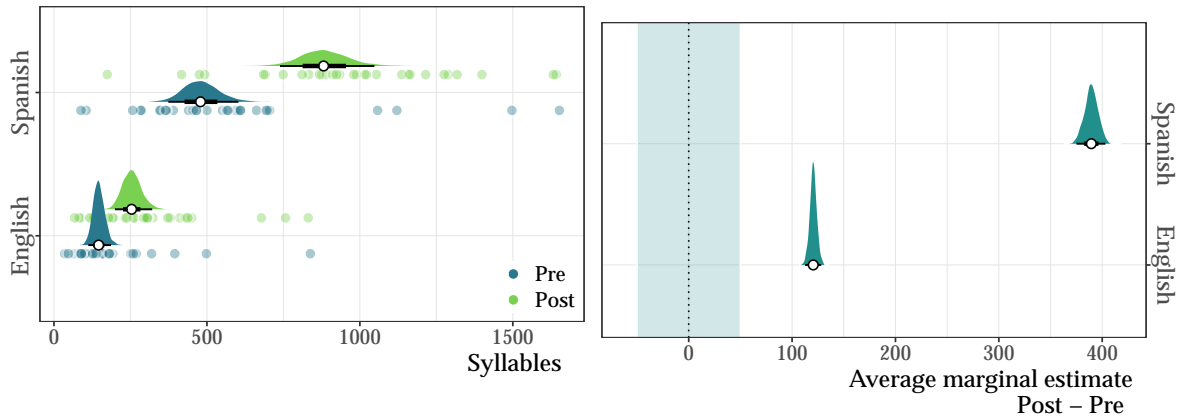
```
avg_comparisons(
  model = mod_sr_n_syllables,
  variables = "phase",
  by = "language"
) |>
  posterior_draws() |>
  ggplot() +
  aes(x = draw, y = language) +
  geom_rect(
    inherit.aes = F,
    data = tibble(xmin = -50, xmax = 50),
    aes(xmin = xmin, xmax = xmax, ymin = -Inf, ymax = Inf),
    fill = my_col(b = 0.5, n = 1, a = 0.2), color = "white",
  ) +
```

```
geom_vline(xintercept = 0, lty = 3) +
stat_halfeye(
  aes(alpha = after_stat(abs(x) < 50)),
  slab_fill = my_col(b = 0.5, n = 1), slab_color = "white",
  show.legend = F, pch = 21, point_fill = "white"
) +
scale_alpha_manual(values = c(1, 0.4)) +
scale_y_discrete(position = "right") +
coord_cartesian(xlim = c(-60, NA)) +
labs(x = "Average marginal estimate\nPost - Pre", y = NULL) +
theme(
  axis.text.y.right = element_text(angle = -90, hjust = 0.5, size = 16),
  axis.text.x = element_text(size = 14),
  axis.title.x = element_text(size = 16)
)
```


(a) Number of syllables produced as a function of language (English, Spanish) and phase (pre, post).

(b) Average marginal estimate of the difference between 'pre' and 'post' for each language.

Figure 1: Syllables produced

```
bind_rows(
n_syllables_post |>
  group_by(language, phase) |>
  median_hdi(value) |>
  mutate(across(value:.upper, \(x) specify_decimal(x, k = 2))) |>
  transmute(
    Language = glue("{language}: {phase}"),
    Estimate = value,
```

```
    HDI = glue("[{.lower}, {.upper}]")
  ) ,

avg_comparisons(
  model = mod_sr_n_syllables,
  variables = "phase",
  by = "language"
) |>
  as_tibble() |>
  mutate(
    contrast = str_remove_all(contrast, "mean\\(|\\)"),
    contrast = str_to_title(contrast)
  ) |>
  mutate(across(estimate:conf.high, \(x) specify_decimal(x, k = 2))) |>
  transmute(
    Language = glue("{language}: {contrast}"),
    Estimate = estimate,
    HDI = glue("[{conf.low}, {conf.high}]")
  )
) |>
  tt(width = 0.75) |>
  style_tt(j = 2:3, align = "r") |>
  group_tt(i = list("Average marginal estimate" = 5)) |>
  style_tt(i = 0, line = "t", line_color = "black", line_width = 0.1) |>
  style_tt(i = 4, line = "b", line_color = "black", line_width = 0.05) |>
  style_tt(i = 7, line = "b", line_color = "black", line_width = 0.1)
```

Number of pauses

```
get_prior(
  formula = bf_raw,
  family = poisson(link = "log"),
  data = filter(dat_z, metric == "n_pauses")
) |>
  select(prior, class, coef) |>
  knitr::kable()
```

Table 2: Number of syllables

Language	Estimate	HDI
English: Pre	145.99	[109.85, 183.60]
English: Post	253.74	[196.62, 317.39]
Spanish: Pre	478.91	[370.56, 596.27]
Spanish: Post	881.32	[739.35, 1047.89]
Average marginal estimate		
English: Post - Pre	120.46	[112.57, 128.22]
Spanish: Post - Pre	389.46	[375.11, 403.13]

prior	class	coef
	b	
	b	phasepost:languageEnglish
	b	phasepost:languageSpanish
	b	phasepre:languageEnglish
	b	phasepre:languageSpanish
lkj(1)	cor	
	cor	
student_t(3, 0, 2.5)	sd	
	sd	
	sd	Intercept
	sd	phasepost:languageEnglish
	sd	phasepost:languageSpanish
	sd	phasepre:languageEnglish
	sd	phasepre:languageSpanish

```
n_pauses_prior <- c(
  prior(normal(4, 1), class = "b"),
  prior(lkj(1), class = "cor"),
  prior(exponential(1), class = "sd")
)

mod_sr_n_pauses <- brm(
  formula = bf_raw,
  family = poisson(link = "log"),
```

```r
    prior = n_syllables_prior,
    sample_prior = "no",
    chains = 4, iter = 2000, warmup = 1000,
    cores = 4, threads = threading(2),
    backend = "cmdstanr",
    control = list(adapt_delta = 0.99, max_treedepth = 12),
    data = filter(dat_z, metric == "n_pauses"),
    file = here("models", "mod_sr_n_pauses")
)
```

```r
n_pauses_post <- as_tibble(mod_sr_n_pauses) |>
  transmute(
    English_pre = exp(`b_phasepre:languageEnglish`),
    English_post = exp(`b_phasepost:languageEnglish`),
    Spanish_pre = exp(`b_phasepre:languageSpanish`),
    Spanish_post = exp(`b_phasepost:languageSpanish`)
  ) |>
  pivot_longer(cols = everything()) |>
  separate(name, into = c("language", "phase"), sep = "_") |>
  mutate(
    phase = str_to_title(phase),
    phase = fct_relevel(phase, "Pre")
  )

n_pauses_post |>
  ggplot() +
  aes(y = language, x = value, group = phase) +
  geom_point(
    data = filter(dat_z, metric == "n_pauses") |>
      mutate(phase = str_to_title(phase), phase = fct_relevel(phase, "Pre")),
    mapping = aes(y = language, x = val, fill = phase),
    position = ggpp::position_dodgenudge(
      x = 0, direction = "split", width = 0.5
    ),
    pch = 21, alpha = 0.4, size = 3, color = "white", stroke = 1
  ) +
  stat_halfeye(
    aes(fill = phase),
    position = ggpp::position_dodgenudge(
      y = 0.06, direction = "split", width = 0.5),
    pch = 21, point_fill = "white", show.legend = F
    ) +
```

```r
    scale_fill_viridis_d(name = NULL, option = "D", begin = 0.4, end = 0.8) +
    scale_color_viridis_d(name = NULL, option = "D", begin = 0.4, end = 0.8) +
    labs(y = NULL, x = "Pauses\n") +
    guides(fill = guide_legend(override.aes = list(alpha = 1, size = 3))) +
    theme(
      legend.position = "inside", legend.justification.inside = c(1, 0),
      legend.background = element_blank(),
      axis.text.y = element_text(angle = 90, hjust = 0.5, size = 16),
      axis.text.x = element_text(size = 14),
      axis.title.x = element_text(size = 16),
      legend.text = element_text(size = 14)
    )

avg_comparisons(
  model = mod_sr_n_pauses,
  variables = "phase",
  by = "language"
) |>
  posterior_draws() |>
  ggplot() +
  aes(x = draw, y = language) +
  geom_rect(
    inherit.aes = F,
    data = tibble(xmin = -10, xmax = 10),
    aes(xmin = xmin, xmax = xmax, ymin = -Inf, ymax = Inf),
    fill = my_col(b = 0.5, n = 1, a = 0.2), color = "white",
  ) +
  geom_vline(xintercept = 0, lty = 3) +
  stat_halfeye(
    aes(alpha = after_stat(abs(x) < 10)),
    slab_fill = my_col(b = 0.5, n = 1), slab_color = "white",
    show.legend = F, pch = 21, point_fill = "white"
  ) +
  scale_alpha_manual(values = c(1, 0.4)) +
  scale_y_discrete(position = "right") +
  labs(x = "Average marginal estimate\nPost - Pre", y = NULL) +
  theme(
    axis.text.y.right = element_text(angle = -90, hjust = 0.5, size = 16),
    axis.text.x = element_text(size = 14),
    axis.title.x = element_text(size = 16)
  )
```
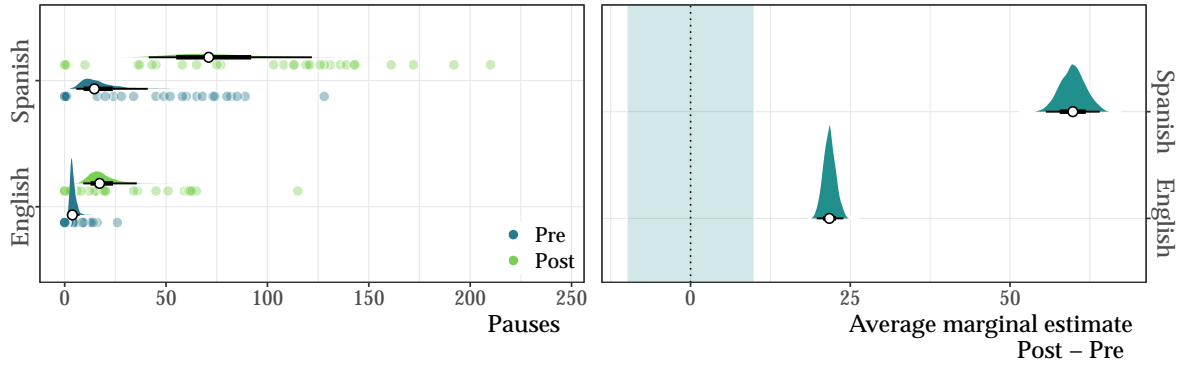
(a) Number of pauses produced as a function of language (English, Spanish) and phase (pre, post).

(b) Average marginal estimate of the difference between 'pre' and 'post' pauses for each language.

Figure 2: Pauses produced

```
bind_rows(
  n_pauses_post |>
    group_by(language, phase) |>
    median_hdi(value) |>
    slice(-4) |>
    mutate(across(value:.upper, \(x) specify_decimal(x, k = 2))) |>
    transmute(
      Language = glue("{language}: {phase}"),
      Estimate = value,
      HDI = glue("[{.lower}, {.upper}]")
    ),

  avg_comparisons(
    model = mod_sr_n_pauses,
    variables = "phase",
    by = "language"
  ) |>
    as_tibble() |>
    mutate(
      contrast = str_remove_all(contrast, "mean\\(|\\)"),
      contrast = str_to_title(contrast)
    ) |>
    mutate(across(estimate:conf.high, \(x) specify_decimal(x, k = 2))) |>
    transmute(
      Language = glue("{language}: {contrast}"),
      Estimate = estimate,
```

Table 4: Number of pauses

Language	Estimate	HDI
English: Pre	3.79	[1.78, 6.80]
English: Post	17.23	[7.57, 31.66]
Spanish: Pre	14.62	[3.64, 32.20]
Spanish: Post	71.02	[38.46, 115.61]
Average marginal estimate		
English: Post - Pre	21.73	[19.74, 23.91]
Spanish: Post - Pre	59.84	[55.62, 64.04]

```
    HDI = glue("[{conf.low}, {conf.high}]")
  )
) |>
tt(width = 0.75) |>
style_tt(j = 2:3, align = "r") |>
group_tt(i = list("Average marginal estimate" = 5)) |>
style_tt(i = 0, line = "t", line_color = "black", line_width = 0.1) |>
style_tt(i = 4, line = "b", line_color = "black", line_width = 0.05) |>
style_tt(i = 7, line = "b", line_color = "black", line_width = 0.1)
```

Articiulation rate

```
get_prior(
  formula = bf_z,
  family = gaussian(),
  data = filter(dat_z, metric == "ar")
) |>
  select(prior, class, coef) |>
  knitr::kable()
```

prior	class	coef
	b	

prior	class	coef
	b	phasepost:languageEnglish
	b	phasepost:languageSpanish
	b	phasepre:languageEnglish
	b	phasepre:languageSpanish
lkj(1)	cor	
	cor	
student_t(3, 0, 2.5)	sd	
	sd	
	sd	Intercept
	sd	phasepost:languageEnglish
	sd	phasepost:languageSpanish
	sd	phasepre:languageEnglish
	sd	phasepre:languageSpanish
student_t(3, 0, 2.5)	sigma	

```r
ar_prior <- c(
  prior(normal(0, 2.5), class = "b"),
  prior(lkj(1), class = "cor"),
  prior(exponential(2), class = "sd"),
  prior(exponential(2), class = "sigma")
)

mod_sr_ar <- brm(
  formula = bf_z,
  family = gaussian(),
  prior = ar_prior,
  sample_prior = "no",
  chains = 4, iter = 2000, warmup = 1000,
  cores = 4, threads = threading(2),
  backend = "cmdstanr",
  data = filter(dat_z, metric == "ar"),
  file = here("models", "mod_sr_ar")
)
```

```r
ar_post <- as_tibble(mod_sr_ar) |>
  transmute(
    English_pre = `b_phasepre:languageEnglish`,
    English_post = `b_phasepost:languageEnglish`,
    Spanish_pre = `b_phasepre:languageSpanish`,
    Spanish_post = `b_phasepost:languageSpanish`
```

```
  ) |>
  pivot_longer(cols = everything()) |>
  separate(name, into = c("language", "phase"), sep = "_") |>
  mutate(
    phase = str_to_title(phase),
    phase = fct_relevel(phase, "Pre")
  )

ar_post |>
  ggplot() +
  aes(y = language, x = value, group = phase) +
  geom_point(
    data = filter(dat_z, metric == "ar") |>
      mutate(phase = str_to_title(phase), phase = fct_relevel(phase, "Pre")),
    mapping = aes(y = language, x = val_z, fill = phase),
    position = ggpp::position_dodgenudge(
      x = 0, direction = "split", width = 0.5
    ),
    pch = 21, alpha = 0.4, size = 3, color = "white", stroke = 1
  ) +
  stat_halfeye(
    aes(fill = phase),
    position = ggpp::position_dodgenudge(
      y = 0.06, direction = "split", width = 0.5),
    pch = 21, point_fill = "white", show.legend = F
    ) +
  scale_fill_viridis_d(name = NULL, option = "D", begin = 0.4, end = 0.8) +
  scale_color_viridis_d(name = NULL, option = "D", begin = 0.4, end = 0.8) +
  labs(y = NULL, x = "Articulation rate") +
  guides(fill = guide_legend(override.aes = list(alpha = 1, size = 3))) +
  theme(
    legend.position = "inside", legend.justification.inside = c(1, 0),
    legend.background = element_blank(),
    axis.text.y = element_text(angle = 90, hjust = 0.5, size = 16),
    axis.text.x = element_text(size = 14),
    axis.title.x = element_text(size = 16),
    legend.text = element_text(size = 14)
  )

avg_comparisons(
  model = mod_sr_ar,
  variables = "phase",
```
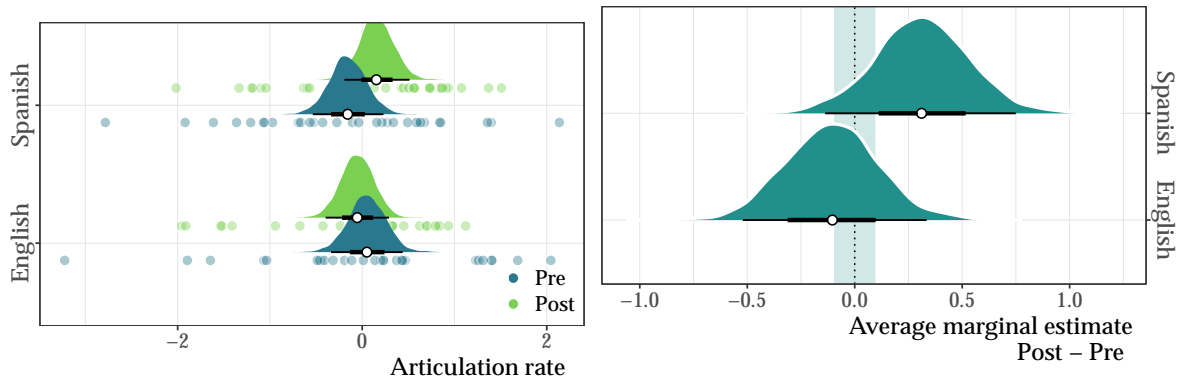
```
  by = "language"
) |>
  posterior_draws() |>
  ggplot() +
  aes(x = draw, y = language) +
  geom_rect(
    inherit.aes = F,
    data = tibble(xmin = -0.1, xmax = 0.1),
    aes(xmin = xmin, xmax = xmax, ymin = -Inf, ymax = Inf),
    fill = my_col(b = 0.5, n = 1, a = 0.2), color = "white",
  ) +
  geom_vline(xintercept = 0, lty = 3) +
  stat_halfeye(
    aes(alpha = after_stat(abs(x) < 10)),
    slab_fill = my_col(b = 0.5, n = 1), slab_color = "white",
    show.legend = F, pch = 21, point_fill = "white"
  ) +
  scale_alpha_manual(values = c(1, 0.4)) +
  scale_y_discrete(position = "right") +
  labs(x = "Average marginal estimate\nPost - Pre", y = NULL) +
  theme(
    axis.text.y.right = element_text(angle = -90, hjust = 0.5, size = 16),
    axis.text.x = element_text(size = 14),
    axis.title.x = element_text(size = 16)
  )
```


(a) Average articulation rate as a function of language (English, Spanish) and phase (pre, post).

(b) Average marginal estimate of the difference between 'pre' and 'post' articulation rate for each language.

Figure 3: Articulation rate

```r
bind_rows(
  ar_post |>
    group_by(language, phase) |>
    median_hdi(value) |>
    mutate(across(value:.upper, \(x) specify_decimal(x, k = 2))) |>
    transmute(
      Language = glue("{language}: {phase}"),
      Estimate = value,
      HDI = glue("[{.lower}, {.upper}]")
    ),

  avg_comparisons(
    model = mod_sr_ar,
    variables = "phase",
    by = "language"
  ) |>
    as_tibble() |>
    mutate(
      contrast = str_remove_all(contrast, "mean\\(|\\)"),
      contrast = str_to_title(contrast)
    ) |>
    mutate(across(estimate:conf.high, \(x) specify_decimal(x, k = 2))) |>
    transmute(
      Language = glue("{language}: {contrast}"),
      Estimate = estimate,
      HDI = glue("[{conf.low}, {conf.high}]")
    )
) |>
  tt(width = 0.75) |>
  style_tt(j = 2:3, align = "r") |>
  group_tt(i = list("Average marginal estimate" = 5)) |>
  style_tt(i = 0, line = "t", line_color = "black", line_width = 0.1) |>
  style_tt(i = 4, line = "b", line_color = "black", line_width = 0.05) |>
  style_tt(i = 7, line = "b", line_color = "black", line_width = 0.1)
```

Average syllable duration

```r
get_prior(
  formula = bf_z,
```

Table 6: Articulation rate

Language	Estimate	HDI
English: Pre	0.05	[-0.34, 0.44]
English: Post	-0.05	[-0.39, 0.29]
Spanish: Pre	-0.16	[-0.53, 0.23]
Spanish: Post	0.15	[-0.19, 0.52]
Average marginal estimate		
English: Post - Pre	-0.10	[-0.52, 0.33]
Spanish: Post - Pre	0.31	[-0.14, 0.75]

```
  family = gaussian(),
  data = filter(dat_z, metric == "asd")
) |>
  select(prior, class, coef) |>
  knitr::kable()
```

prior	class	coef
	b	
	b	phasepost:languageEnglish
	b	phasepost:languageSpanish
	b	phasepre:languageEnglish
	b	phasepre:languageSpanish
lkj(1)	cor	
	cor	
student_t(3, 0, 2.5)	sd	
	sd	
	sd	Intercept
	sd	phasepost:languageEnglish
	sd	phasepost:languageSpanish
	sd	phasepre:languageEnglish
	sd	phasepre:languageSpanish
student_t(3, 0, 2.5)	sigma	

```r
asd_prior <- c(
  prior(normal(0, 2.5), class = "b"),
  prior(lkj(1), class = "cor"),
  prior(exponential(2), class = "sd"),
  prior(exponential(2), class = "sigma")
)

mod_sr_asd <- brm(
  formula = bf_z,
  family = gaussian(),
  prior = asd_prior,
  sample_prior = "no",
  chains = 4, iter = 2000, warmup = 1000,
  cores = 4, threads = threading(2),
  backend = "cmdstanr",
  control = list(adapt_delta = 0.99),
  data = filter(dat_z, metric == "asd"),
  file = here("models", "mod_sr_asd")
)
```

```r
asd_post <- as_tibble(mod_sr_asd) |>
  transmute(
    English_pre = `b_phasepre:languageEnglish`,
    English_post = `b_phasepost:languageEnglish`,
    Spanish_pre = `b_phasepre:languageSpanish`,
    Spanish_post = `b_phasepost:languageSpanish`
  ) |>
  pivot_longer(cols = everything()) |>
  separate(name, into = c("language", "phase"), sep = "_") |>
  mutate(
    phase = str_to_title(phase),
    phase = fct_relevel(phase, "Pre")
  )

asd_post |>
  ggplot() +
  aes(y = language, x = value, group = phase) +
  geom_point(
    data = filter(dat_z, metric == "asd") |>
      mutate(phase = str_to_title(phase), phase = fct_relevel(phase, "Pre")),
    mapping = aes(y = language, x = val_z, fill = phase),
    position = ggpp::position_dodgenudge(
```

```
      x = 0, direction = "split", width = 0.5
    ),
    pch = 21, alpha = 0.4, size = 3, color = "white", stroke = 1
  ) +
  stat_halfeye(
    aes(fill = phase),
    position = ggpp::position_dodgenudge(
      y = 0.06, direction = "split", width = 0.5),
    pch = 21, point_fill = "white", show.legend = F
    ) +
  scale_fill_viridis_d(name = NULL, option = "D", begin = 0.4, end = 0.8) +
  scale_color_viridis_d(name = NULL, option = "D", begin = 0.4, end = 0.8) +
  labs(y = NULL, x = "Average Syllable Duration") +
  guides(fill = guide_legend(override.aes = list(alpha = 1, size = 3))) +
  theme(
    legend.position = "inside", legend.justification.inside = c(1, 0),
    legend.background = element_blank(),
    axis.text.y = element_text(angle = 90, hjust = 0.5, size = 16),
    axis.text.x = element_text(size = 14),
    axis.title.x = element_text(size = 16),
    legend.text = element_text(size = 14)
  )

avg_comparisons(
  model = mod_sr_asd,
  variables = "phase",
  by = "language"
) |>
  posterior_draws() |>
  ggplot() +
  aes(x = draw, y = language) +
  geom_rect(
    inherit.aes = F,
    data = tibble(xmin = -0.1, xmax = 0.1),
    aes(xmin = xmin, xmax = xmax, ymin = -Inf, ymax = Inf),
    fill = my_col(b = 0.5, n = 1, a = 0.2), color = "white",
  ) +
  geom_vline(xintercept = 0, lty = 3) +
  stat_halfeye(
    aes(alpha = after_stat(abs(x) < 10)),
    slab_fill = my_col(b = 0.5, n = 1), slab_color = "white",
    show.legend = F, pch = 21, point_fill = "white"
```
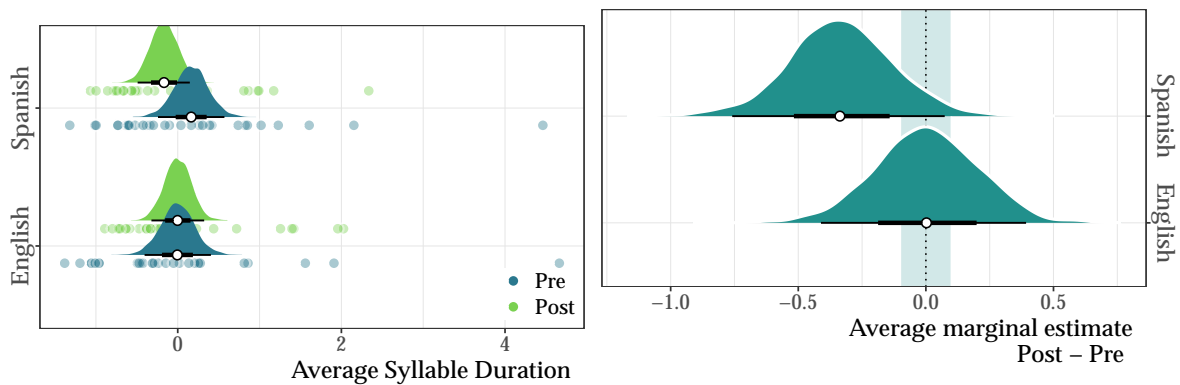
```
) +
scale_alpha_manual(values = c(1, 0.4)) +
scale_y_discrete(position = "right") +
labs(x = "Average marginal estimate\nPost - Pre", y = NULL) +
theme(
  axis.text.y.right = element_text(angle = -90, hjust = 0.5, size = 16),
  axis.text.x = element_text(size = 14),
  axis.title.x = element_text(size = 16)
)
```


(a) Average syllable duration as a function of language (English, Spanish) and phase (pre, post).

(b) Average marginal estimate of the difference between 'pre' and 'post' articulation rate for each language.

Figure 4: Average syllable duration

```
bind_rows(
  asd_post |>
    group_by(language, phase) |>
    median_hdi(value) |>
    mutate(across(value:.upper, \(x) specify_decimal(x, k = 2))) |>
    transmute(
      Language = glue("{language}: {phase}"),
      Estimate = value,
      HDI = glue("[{.lower}, {.upper}]")
    ),

  avg_comparisons(
    model = mod_sr_asd,
    variables = "phase",
    by = "language"
```

Table 8: Articulation rate

Language	Estimate	HDI
English: Pre	-0.01	[-0.41, 0.40]
English: Post	0.00	[-0.34, 0.30]
Spanish: Pre	0.16	[-0.24, 0.57]
Spanish: Post	-0.17	[-0.50, 0.14]
Average marginal estimate		
English: Post - Pre	0.00	[-0.41, 0.39]
Spanish: Post - Pre	-0.34	[-0.76, 0.07]

```
) |>
  as_tibble() |>
  mutate(
    contrast = str_remove_all(contrast, "mean\\(|\\)"),
    contrast = str_to_title(contrast)
  ) |>
  mutate(across(estimate:conf.high, \(x) specify_decimal(x, k = 2))) |>
  transmute(
    Language = glue("{language}: {contrast}"),
    Estimate = estimate,
    HDI = glue("[{conf.low}, {conf.high}]")
  )
) |>
tt(width = 0.75) |>
style_tt(j = 2:3, align = "r") |>
group_tt(i = list("Average marginal estimate" = 5)) |>
style_tt(i = 0, line = "t", line_color = "black", line_width = 0.1) |>
style_tt(i = 4, line = "b", line_color = "black", line_width = 0.05) |>
style_tt(i = 7, line = "b", line_color = "black", line_width = 0.1)
```

Chang, Charles B. 2012. "Rapid and Multifaceted Effects of Second-Language Learning on First-Language Speech Production." *Journal of Phonetics* 40 (2): 249–68.

———. 2013. "A Novelty Effect in Phonetic Drift of the Native Language." *Journal of Phonetics* 41 (6): 520–33.

Flege, James Emil, and Ocke-Schwen Bohn. 2021. "The Revised Speech Learning Model (SLM-r)." In *Second Language Speech Learning: Theoretical and Empirical Progress*, edited by Ra-

tree Wayland, 3–83. Cambridge University Press. https://doi.org/10.1017/9781108886901.
002.

Lang, Benjamin, and Lisa Davidson. 2019. "Effects of Exposure and Vowel Space Distribution
on Phonetic Drift: Evidence from American English Learners of French." *Language and
Speech* 62 (1): 30–60.

Tobin, Stephen J, Hosung Nam, and Carol A Fowler. 2017. "Phonetic Drift in Spanish-English
Bilinguals: Experiment and a Self-Organizing Model." *Journal of Phonetics* 65: 45–59.

Van Leussen, Jan-Willem, and Paola Escudero. 2015. "Learning to Perceive and Recognize
a Second Language: The L2LP Model Revised." *Frontiers in Psychology* 6: 6–12. https:
//doi.org/10.3389/fpsyg.2015.01000.