Solving cross-language variable equivalence in bilingual clause combining

Dora LaCasse & Rena Torres Cacoullos
University of Montana & Pennsylvania State University

15th Biennial High Desert Linguistics Society Conference
University of New Mexico
Albuquerque, November 11-13, 2022
Bilingual clause combining: Language of the conjunction?

- Code-switching (CS) = going back and forth between languages

and,  
they’re just letting their money build up,  
porque no lo pueden pescar porque  
ya está adentro.

‘and,  
they’re just letting their money build up,  
because they can’t get at them because  
they’re already inside.’

[20, 33:04–33:08]
Bilingual clause combining: Language of the conjunction?

- CS constrained by cross-language equivalence / congruence
  (e.g., Deuchar 2005: 255; Lipski 1978: 257–258; Muysken 2015: 259)

- Equivalence Constraint
  - operationalizable for bilingual speech corpus
  - does not assume that CS constraints derive from
    - general syntactic principles of monolingual grammar (e.g. MacSwan 2020)
    - the syntax of one of the languages, the Matrix Language (e.g., Myers-Scotton 1993)

- CS is avoided at points of word placement incompatibility between the two languages
  - CS occurs at the kinds of syntactic boundaries that are present in both languages
Bilingual clause combining: Language of the conjunction?

• Equivalence Constraint:
  • Code-switching (CS) = concatenated multiword strings of alternating languages, each of which is grammatical in its respective language of origin
    • Boundary between adjacent strings occurs where elements are placed the same way in both languages


they’re just letting their money build up, porque no lo pueden pescar

‘they’re just letting their money build up, because they can’t get at them’

finite verb_{LANG X} + conjunction + finite verb_{LANG Y}
Bilingual clause combining: Language of the conjunction?

• Points of Variable Equivalence
  • where equivalence at the boundary between languages is not consistent due to internal variability within one or both of the languages
Equivalence

- Main-and-adverbial clause:
  
  finite verb + causal conjunction + finite verb

<table>
<thead>
<tr>
<th>SPANISH</th>
<th>ENGLISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>porque</td>
<td>(be)cause</td>
</tr>
<tr>
<td>presence 100%</td>
<td>presence 100%</td>
</tr>
</tbody>
</table>
Variable equivalence

• Main-and-complement clause:

finite verb + complementizer + finite verb

SPANISH

que
presence 100%

(Silva-Corvalán 1994:137; Torres Cacoullos et al. 2017:81), excepting formulaic expressions, e.g., well-wishing espero... 'I hope' (cf. Mazzola et al. 2022:24; Rodríguez Ricelli, 2018: 323–327)

ENGLISH

that
presence vs. absence

Bilingual clause combining: Language of the complementizer?

(1) Main + *que* + Complement

\[ \text{me dijeron que,} \]
\[ \text{I was gonna run the !two mile?} \]

‘they told me that, \text{I was gonna run the !two mile?}’

[22, 11:07-11:09]

(2) Main + *that* + Complement

\[ \text{.. ella me dijo,} \]
\[ \text{.. that she’d rather go to ~Nancy’s.} \]

‘.. she told me, \text{.. that she’d rather go to ~Nancy’s.}’

[31, 21:53-21:55]

(3) Main + *Ø* + Complement

\[ \text{I guess Ø no me podía defender.} \]

'I guess Ø I couldn't defend myself'

[06, 6:58-7:00]
Solving cross-language variable equivalence in bilingual clause combining

• Spanish *que* is overwhelmingly selected over English *that*
  • but not Spanish *porque* over English *(be)cause*

• Variable equivalence hypothesis for CS:
  where cross-language equivalence is not consistent due to variability within one of the languages, bilinguals restore equivalence by choosing the form that is:
  • 1. more quantitatively available
  • 2. more automatic
  in their combined linguistic experience, accounting for their use of both languages
Data: Speech community and corpus

New Mexico Spanish-English Bilingual (NMSEB) corpus

https://nmcode-switching.la.psu.edu/
NMSEB corpus participants ($N=40$) = regularly use both languages
- women ($n=23$), men ($n=17$)
- born 1922-1993
- mineworkers, ranchers, teachers, service employees
- mostly rural areas (72%)

- 31 sociolinguistic interview recordings
  (Labov 1984)
- 29 hours – 300,000 words
Data

• Main-and-complement clause
  finite verb + complementizer + finite verb
  • $N = 1,136$

  (Steuck 2016: 77–80 on data extraction protocols and exclusions)

• Main-and-adverbial clause
  finite verb + causal conjunction + finite verb
  • $N = 1,061$

  (porque and (be)cause clauses with an associated main clause)
• Unilingual Spanish
  
  yo pensé que estaba muy alto.  
  'I thought that it was very high.'
  
• Unilingual English
  a. ...(0.6) but it's good that they're both involved.
  [8 intervening lines]
  b. I think I know the mother and the father,
  c. and yeah,
  d. they're both good people.
  
  [21, 43:31-43:47]
CS rate

• Data: main-and-complement clause

Bilinguals' main-and-complement clauses according to language ($N=1,136$)

- CS between clauses: 6%
- Unilingual Spanish: 43%
- Unilingual English: 41%
- Other:

I think Ø I know the mother and the father,

[21, 43:43-43:46]

yo pensé que estaba muy alto.

[31, 52:11–52:12]
CS rate

• Data: main-and-complement clause

Bilinguals' main-and-complement clauses according to language ($N=1,136$)

$I \text{ think } \varnothing \text{ he had another one } \text{ allá también,}$

[23, 23:45–23:47]
CS rate

- Data: main-and-complement clause

Bilinguals' main-and-complement clauses according to language ($N=1,136$)

<table>
<thead>
<tr>
<th></th>
<th>CS between clauses</th>
<th>Unilingual Spanish</th>
<th>Unilingual English</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

finite verb$_{\text{LANG } X}$ + complementizer + finite verb$_{\text{LANG } Y}$

*so I told them,*

que iban a salir en el Sun.

[22, 17:05–17:08]
CS rate

• Data: main-and-adverbial clause

Bilinguals' main-and-adverbial clauses according to language ($N=1,061$)
CS rate

• Data: main-and-adverbial clause

Bilinguals' main-and-adverbial clauses according to language ($N=1,061$)

finite verb $^{\text{LANG X}}$ + causal conjunction + finite verb $^{\text{LANG Y}}$

they’re just letting their money build up, porque no lo pueden pescar

[06, 1:08:37-1:08:39]
CS rate

Bilinguals' main-and-complement clauses ($N=1,136$) vs. main-and-adverbial clauses ($N=1,061$) according to language
Variable equivalence

• Equivalence site
  • Main-and-adverbial clause
    • CS rate: higher

• Variable Equivalence site
  • Main-and-complement clause
    • CS rate: lower
Bilingual clause combining: Language of the complementizer?

(1) Main + *que* + Complement

\[
\text{me dijeron } \textit{que}, \\
\text{I was gonna run the } \text{!two mile?} \\
\]

‘they told me \textit{that}, \\
\text{I was gonna run the } \text{!two mile?’}

[22, 11:07-11:09]

(2) Main + *that* + Complement

\[
.. \text{ella me dijo,} \\
.. \text{that she’d rather go to } \sim\text{Nancy’s.} \\
\]

‘.. she told me, \\
.. \text{that she’d rather go to } \sim\text{Nancy’s.’}

[31, 21:53-21:55]

(3) Main + *Ø* + Complement

\[
\text{I guess } \text{Ø no me podía defender.} \\
\]

‘I guess Ø I couldn’t defend myself’

[06, 6:58 -7:00]
Bilingual clause combining: Language of the complementizer?

• = language of the main verb
  • Matrix Language Frame
    
    *me dijeron que,*  \quad \textit{they told me that,}  
    *I was gonna run the !two mile?*  \quad \textit{I was gonna run the !two mile?}

  

• = language of the subordinate verb
  • Formal syntax: “no CS-specific constraints”
    
    (MacSwan 2020:94)

  • “Functional head”, “phase”
    
    (e.g., Belazi et al. 1994, López et al. 2017)

  *.. ella me dijo,*  \quad \textit{.. she told me,}  
  *.. that she’d rather go to \sim Nancy’s.*  \quad \textit{.. that she’d rather go to \sim Nancy’s.}
Results: Language of the complementizer = Spanish *que*

Use of complementizers in code-switched main-and-complement clause sentences (N=66)

- Bilinguals' CS between clauses
  - THAT: 5%
  - Ø: 6%
  - QUE: 86%
Results: Language of the complementizer = Spanish *que*

Language of the complementizer according to CS direction \((n=60)\)

<table>
<thead>
<tr>
<th>Language to Language</th>
<th>THAT</th>
<th>QUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>English to Spanish</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Spanish to English</td>
<td></td>
<td>94%</td>
</tr>
</tbody>
</table>

English to Spanish

*so I told them,*

*que iban a salir*

*en el Sun.*

Spanish to English

*me dijeron que,*

*I was gonna run*

*the !two mile?*

'so I told them,

*that they were going*

*to be in the Sun.*'

[22, 17:05–17:08]

‘they told me *that,*

*I was gonna run*

*the !two mile?*

[22, 11:07-11:09]
Bilingual clause combining: Language of the complementizer?

• = language of the main verb
  • Matrix Language Frame
    
    *me dijeron que, I was gonna run the two mile?*  
    *I was gonna run the two mile?*
    
  (e.g., Myers-Scotton 1993, 2002, etc.)

• = language of the subordinate verb
  • Formal syntax: “no CS-specific constraints”
    
    (MacSwan 2020:94)

  • Functional head, phase
    
    *.. ella me dijo, .. that she'd rather go to ~Nancy’s.*  
    *.. she told me, .. that she'd rather go to ~Nancy’s.*

  (e.g., Belazi et al. 1994, López et al. 2017)
Bilingual clause combining: Why Spanish complementizer *que*?

- Spanish as dominant or Matrix Language?
  - Main-and-complement and main-and-adverbial clause distributions by language
    - even between English and Spanish ✓

![Graph showing distribution of main-and-complement and main-and-adverbial clauses by language]

Bilingual clause combining: Why Spanish complementizer *que*?

• Spanish as dominant or Matrix Language?
  • Main-and-complement and main-and-adverbial clause distributions by language
    • even between English and Spanish ✓
  • Overall corpus
    • distribution of clauses by language is even ✓
Bilingual clause combining: Why Spanish complementizer *que*?

NMSEB corpus
Even distribution of clauses by language
\( N = 36,011 \)

(Torres Cacoullos & Travis 2018/2020)
Bilingual clause combining: Why Spanish complementizer *que*?

• **Spanish as dominant or Matrix Language?**
  • Main-and-complement and main-and-adverbial clauses distributions by language
    • even between English and Spanish ✓
  • Overall corpus
    • distribution of clauses by language is even ✓
    • intra-sentential CS is bidirectional ✓
Bilingual clause combining: Why Spanish complementizer *que*?

- **Both**
  (e.g., bring -- that name, porque no había ~Roberts, todavía, in the family. [04, 15:59])

- **Spanish**
  (e.g., ~Alma was more flaca que la mama. [04, 1:04:18])

- **English**
  (e.g., pues estaban asina como caddy corner across the street. [04, 1:02:12])

---

**NMSEB corpus**

**Bidirectionality of CS**

*N = 2,068*

(Trawick, 2022, Table 6.12)
Bilingual clause combining: Why Spanish complementizer *que*?

Spanish as dominant or Matrix Language?

<table>
<thead>
<tr>
<th></th>
<th>CSs</th>
<th>Clauses (overall)</th>
<th>Main-and-complement</th>
<th>Main-and-adverbial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unilingual Spanish</td>
<td>39%</td>
<td>51%</td>
<td>49%</td>
<td>49%</td>
</tr>
<tr>
<td>Unilingual English</td>
<td>36%</td>
<td>51%</td>
<td>43%</td>
<td>35%</td>
</tr>
</tbody>
</table>
Bilingual clause combining: Why Spanish complementizer *que*?

- Why Spanish complementizer *que*?
  - Spanish as dominant or Matrix Language? X

Variable Equivalence hypothesis

• Equivalence Constraint
  • CS is avoided at points of word placement incompatibility between the two languages
    • occurs at the kinds of syntactic boundaries that are present in both languages

• Variable Equivalence points
  • syntactic boundaries that occur variably within one or both of the languages, so that cross-language equivalence is not consistent
    (Torres Cacoullos & Poplack 2016)
Variable Equivalence

- Main-and-complement clause:
  finite verb + complementizer + finite verb

SPANISH

que presence
100% (Silva-Corvalán 1994:137; Torres Cacoullos et al. 2017:81), excepting formulaic expressions, e.g., well-wishing espero... 'I hope' (cf. Mazzola et al. 2022:24; Rodríguez Ricelli, 2018: 323–327)

ENGLISH

that presence vs. absence (e.g., Shank et al. 2016: 202-213, Tagliamonte and Smith 2005:299-301, Torres Cacoullos and Walker 2009:19-32)
Variable Equivalence hypothesis

• Bilinguals deal with cross-language variable equivalence by choosing the more quantitatively available and more automatic variant form.
  (Torres Cacoullos 2020)

• Conjunction *que* is a more frequent option than *that*
  • which is a minority variant within English
Variable Equivalence hypothesis

Bilinguals’ main-and-complement clauses according to language ($N=1,136$)
que is the more quantitatively available variant in bilinguals’ combined linguistic experience
*que* is the more quantitatively available variant in bilinguals’ combined linguistic experience.
*que* is the more quantitatively available variant in bilinguals’ combined linguistic experience.

---

**Use of complementizers according to main-and-complement clause language** (N=1,017)

- Bilinguals' CS between clauses: 86%
- Bilinguals' unilingual English: 73%
- Bilinguals' unilingual Spanish: 100%

Graph showing the use of complementsizers in different contexts.
Variable Equivalence hypothesis

• Bilinguals deal with cross-language variable equivalence by choosing the more quantitatively available and more automatic variant form (Torres Cacoullos 2020)
  • Conjunction *que* is a more frequent option than *that* ✓
    • which is a minority variant within English
Variable Equivalence hypothesis

• Bilinguals deal with cross-language variable equivalence by choosing the more quantitatively available and more automatic variant form (Torres Cacoullos 2020)

  • Conjunction *que* is a more frequent option than *that* ✓
    • which is a minority variant within English

  • Conjunction *que* is a more discourse-neutral variant form than *that*
    • which is subject to discourse constraints
 que is the more discourse-neutral variant in bilinguals’ combined linguistic experience
"que is the more discourse-neutral variant in bilinguals’ combined linguistic experience"

Linguistic conditioning of complementizer *that* presence:
Direction of effect in monolingual varieties and in bilinguals' English

*that* presence:
lexical, structural and discourse constraints
- main clause verb,
- intervening material,
- main clause subject person
- complement clause subject form (lexical vs. pronominal)

*that* presence is favored by
- main clause *know, say, tell* > *guess, think*
- adverbials, clauses, fillers, pauses between the clauses
- main clause non 1sg > *I*
- complement clause full NP subject > pronouns
Variable Equivalence hypothesis

- Bilinguals deal with cross-language variable equivalence by choosing the more quantitatively available and more automatic variant form (Torres Cacoullos 2020)
  - Conjunction *que* is a more frequent option than *that* ✓
    - which is a minority variant within English
  - Conjunction *que* is a more discourse-neutral variant form than *that* ✓
    - which is subject to discourse constraints
Variable Equivalence hypothesis

• Bilinguals deal with cross-language variable equivalence by choosing the more quantitatively available and more automatic variant form (Torres Cacoullos 2020)

  • Conjunction *que* is a more frequent option than *that* ✓
    • which is a minority variant within English

  • Conjunction *que* is a more discourse-neutral variant form than *that* ✓
    • which is subject to discourse constraints

Context-independent *que* is a more likely choice than the probabilistically constrained *that* for restoring equivalence at the boundary of the two languages
Variable equivalence

• Final test: bilingual clause combining at equivalence site
  • main-and-adverbial clause, with Spanish *porque* vs. English *(be)*cause
Equivalence

• Main-and-adverbial clause:
  finite verb + causal conjunction + finite verb

SPANISH

**porque**
presence 100%

ENGLISH

*(be)*cause
presence 100%
Bilingual clause combining: Language of the conjunction?

(1) ... *it was so pretty porque entonces tenían color.* ‘... *it was so pretty because* then they had color.’

(2) *y estaba uno que le decían ~Miguel Bajo,* ‘and there was one that they would call ~Miguel Bajo,
*because he must’ve been ~four foot nine,*’

[06, 1:08:37-1:08:39]

[17, 12:15-12:19]
Bilingual clause combining: Language of the conjunction?

Use of conjunctions in code-switched main-and-causal clause combinations (N=127)
Variable equivalence

Language of the complementizer (left) and causal conjunction (right), according to CS direction
Variable equivalence

Language of the complementizer (left) and causal conjunction (right), according to CS direction

English to Spanish | Spanish to English
---|---

English to Spanish | Spanish to English
---|---

THAT | QUE
---|---

(BE)CAUSE | PORQUE
---|---
Bilinguals choose more frequent and automatic option

- Variable equivalence matters
  - bilinguals treat complement and adverbial clauses differently
  - Spanish *que* is overwhelmingly selected over English *that*, but not Spanish *porque* over English *(be)cause*

- Variable equivalence hypothesis
  - Bilinguals deal with language-internal variability:
    - 1. restore cross-language equivalence
    - 2. choose more frequent and more automatic option
Thank you!

Research supported by
NSF project # 1624966
Code-switching in spontaneous bilingual speech

¡Gracias!