Can two unrelated languages be mixed? Evidence from a new method to investigate codemixing

Jeanine Treffers-Daller (UoR), Zehra Ongun (UoR), Çiçe Çavuşoğlu (Near East University, Northern Cyprus, Cyprus), Valentina Christodoulou (University of Cyprus, Cyprus), Theodosia Demetriou (University of Nicosia, Cyprus), Christiana Themistocleous (UoR), Julia Hofweber (University College London) & Michal Korenar (UoR)

RUEG2021 Dynamics of Language Contact,
21-23 February 2021
Overview

- Why study code-switching?
- Different types of code-switching (Muysken, 2000; 2013)
- Developing a Turkish-English code-switching frequency task (CSFT)
- Preliminary results from a study among three groups of Turkish-English bilinguals
  - To what extent does congruent lexicalization occur in data from unrelated languages?
  - Which background variables predict variability in congruent lexicalization? Cultural affiliation or country of residence?
  - How useful is the CSFT for studying variability in CS patterns?
- Discussion – next steps
Why study code-switching (CS)?

- Linguistics: What are the constraints/rules on CS
- Psycholinguistics: How can bilinguals keep their languages separate AND switch in appropriate settings?
- Sociolinguistics: Which socio-cultural variables covary with the CS patterns we find?
Why study CS among heritage speakers?

- Does CS promote or accelerate language change/attrition in heritage speakers? (Backus, 2005)
- Which types of CS are most likely to be mechanisms for structural change?
  - The most intimate types of CS (congruent lexicalization)?
- Which aspects of language structure are most likely to be affected by CS?
  - Word order?
Different types of CS

(A) insertion

(B) congruent lexicalization

L1

L1/L2

UP

L2

(C) alternation

(D) back-flagging

Muysken, 2013
**Insertion**

Chunks from language B inserted into grammar of language A

We didn’t bring **SCHUHWERK** for hiking.
We didn’t bring **SHOES** for hiking.
(Hofweber et al., 2019)
Alternation

Loosely connected phrases from languages A and B alternate
Ich kann heute nicht kommen BECAUSE I’M ILL.
I can today not come BECAUSE I’M ILL.

„I cannot come today because I am ill.“ (Hofweber et al., 2019)
Backflagging
Insertion of heritage language discourse markers in L2 discourse

a. Maya: It was the only way, Mama.

b. Rosa: ¡Ay!, Maya, you’re taking this too far.
   (English/Spanish; Specker, 2008, p. 114)
<table>
<thead>
<tr>
<th>(A) insertion</th>
<th>(B) congruent lexicalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>L1/L2</td>
</tr>
</tbody>
</table>

**Congruent lexicalisation** co-activation of grammar, semantics and lexicon of languages A and B

Wir haben FRIENDS gemacht mit’m SHOP OWNER.
We have FRIENDS made with th’ SHOP OWNER.
We have made FRIENDS with th’ SHOP OWNER (Hofweber et al., 2019)
## Diagnostic criteria for congruent lexicalization

<table>
<thead>
<tr>
<th>Switch Site</th>
<th>Insertion</th>
<th>Alteration</th>
<th>Congruent Lexicalization</th>
<th>Back-Flagging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major clause boundary</td>
<td></td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Peripheral</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Embedding in discourse</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flagging</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>Dummy word insertion</td>
<td>+</td>
<td></td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Bidirectional code-switching</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Properties</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear equivalence</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Telegraphic mixing</td>
<td></td>
<td>−</td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>Morphological integration</td>
<td>+</td>
<td></td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>Doubling</td>
<td>−</td>
<td>+</td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>Homophonous diamorphs</td>
<td>−</td>
<td>+</td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>Triggering</td>
<td></td>
<td></td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>Mixed collocations</td>
<td>−</td>
<td>+</td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>Self-corrections</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Here no value refers to the non-applicability of a specific feature.*

Muysken, Deuchar and Wang, 2007)
Adaptation of the model (Muysken, 2014)

Key distinction:
- **insertion** (clear matrix language imposing its constraints): Essentially selected items
- **alternation** (involving several languages imposing their constraints): Essentially adjuncts

**Congruent lexicalization:** insertion or alternation under condition of similarity between languages

Questions:

a) Can congruent lexicalization („mixing“) be found in languages that are not related?

b) What does it mean for languages to be related (see Ringbom & Jarvis, 2009)?
Gaining deeper insights into variability in CS

- A corpus will give us authentic, ecologically valid data, but limited information about all possible switches + variability
- Frequency data from small bilingual corpora inconclusive
- Questionnaires with self-reported CS may not tap into real life CS because of stigma or lack of awareness. No distinction made between types of CS (Hofweber et al., 2019)
- Experimental techniques needed (Gullberg, Indefrey & Muysken, 2009; Treffers-Daller, 2021)
- Frequency judgments are indicative of cognitive embedding (Backus, 2015)
- Usage-based approach (Croft, 2000), combined with experimental techniques (Onar Valk, 2014; Backus 2015; Hofweber et al. 2016, 2019 and 2020; Treffers-Daller et al., 2021)
Code-switching frequency task (CSFT)

- Stimuli based on naturalistic examples from CS corpora, but controlled for length in syllables
- Oral presentation through headphones with support on PPT
- Presentation in non-standard varieties (if applicable)
- Inclusion of monolingual stimuli (Ebert & Koronkiewicz, 2018)
- Respondents indicate on a Likert scale how frequently they encounter stimuli in their environment (not grammaticality judgement)
- **Validation evidence**: frequency ratings from the CSFT predicted bilinguals’ use of congruent lexicalization in an email production task in German-English CS (see appendix) (Hofweber et al., 2019)
THE CURRENT PROJECT
Research questions

1. Can speakers of two unrelated languages engage in congruent lexicalization (CLX)?
2. Do CS patterns among three groups of Turkish-English bilinguals differ, in particular in relation to CLX?
   - Turks from the mainland (recent immigrants to UK)
   - Turkish Cypriots (in the UK since 1914)
   - Turkish Cypriots living in Cyprus
3. Which variables explain the variance in CLX?
   - Residence (UK versus Cyprus)
   - Cultural affiliation (Turkish versus Cypriot culture)
4. How successful is CSFT in revealing CS patterns?
Hypotheses

1. CLX category with lowest frequency among Turkish-English bilinguals, because of typological distance

2. Bilinguals with longer tradition of language contact with English more likely to engage in CLX (Muysken, 2000; 2013)
   - Cypriots > Turks (cultural affiliation)

3. Bilinguals living in the UK more likely to engage in CLX
   - UK resident Cypriots > Cyprus-based Cypriots
   - Relative importance of culture/residence?
Participants: adult Turkish-English bilinguals

<table>
<thead>
<tr>
<th></th>
<th>Cyprus-based</th>
<th>UK-based</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cypriots</td>
<td>40</td>
<td>28</td>
<td>68</td>
</tr>
<tr>
<td>Turks</td>
<td>-</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>total</td>
<td>40</td>
<td>58</td>
<td>98</td>
</tr>
</tbody>
</table>

Mean age: 32.21 (sd 9.3); 53 males, 45 females

Two informants from the UK-based Cypriots were removed because of outliers (3SD above mean) on code-switching frequency task.
Background variables by place of residence

<table>
<thead>
<tr>
<th></th>
<th>UK Mean</th>
<th>Cyprus Mean</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>1.43</td>
<td>1.50ns</td>
<td></td>
</tr>
<tr>
<td>education (1-4)</td>
<td>2.79</td>
<td>2.63ns</td>
<td></td>
</tr>
<tr>
<td>English self rating</td>
<td>5.84</td>
<td>6.28*</td>
<td></td>
</tr>
<tr>
<td>Turkish self rating</td>
<td>5.62</td>
<td>6.60**</td>
<td></td>
</tr>
<tr>
<td>Turkish years of use</td>
<td>26.21</td>
<td>33.90**</td>
<td></td>
</tr>
<tr>
<td>English years of use</td>
<td>19.66</td>
<td>28.10**</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>28.90</td>
<td>37.03**</td>
<td></td>
</tr>
</tbody>
</table>

Covariates included in further analyses:
- E and T self ratings
- Years of use of T and E
- age
The CS frequency judgement task

- Authentic examples from literature on Turkish-English CS, translated from Turkish-German or Turkish-Danish CS
- Stimuli (98)
  - 14 per code-switching type from authentic corpora (T->E and E->T)
  - 14 mixed verbal compounds (T->E and E->T)
  - 14 switches between utterances (T->E and E->T)
  - 14 monolingual control sentences (7 T and 7 E)
- Standard Turkish + Cypriot Turkish version
Presentation of the CSFT

- **Presentation**: visual and audio format & randomised
- **Instructions**: “How often do you come across this type of sentence when talking to other Turkish/Cypriot people in your environment?”
- **Visual analogue scale** (Llamas & Watt, 2014)

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ben celebrating yapmiyacağım</td>
<td>never</td>
<td>x</td>
<td>always</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ağıstos is içreç, değil mi?</td>
<td>never</td>
<td>x</td>
<td>always</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example of congruent lexicalization (E>T)

In that film many mistakes were mademiş
“In that film many mistakes were made.”
Reliability and dimensional structure  CSFT

- Cronbach’s alpha = .961
- Factor analysis: two dimensions (rotated factor solution, direct oblimin, suppression of factor loadings smaller than .3)

<table>
<thead>
<tr>
<th>Component</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>monoTURK</td>
<td>.548</td>
<td>.705</td>
</tr>
<tr>
<td>monoENG</td>
<td>.422</td>
<td>.762</td>
</tr>
<tr>
<td>INTER_tot</td>
<td>.601</td>
<td></td>
</tr>
<tr>
<td>INS_tot</td>
<td>.902</td>
<td></td>
</tr>
<tr>
<td>BFL_tot</td>
<td>.889</td>
<td></td>
</tr>
<tr>
<td>ALT_tot</td>
<td>.897</td>
<td></td>
</tr>
<tr>
<td>CLX_tot</td>
<td>.830</td>
<td></td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

Monolingual sentences perceived to be a separate dimension

All CS types perceived as part of one dimension
RESULTS
How often do you hear sentences such as X in your environment?
Differences in frequency of monolingual sentences and total CS, by country of residence

(*) F(1) = 3.74, p = .057, \( \eta^p = .04 \)

Respondents in the UK hear more monolingual English sentences than respondents in Cyprus. Important validation evidence for the task.

Total CS significantly lower than monolingual sentences (Friedman, \( Z = -7.88, p < .001 \))
Congruent lexicalization E→T and T→E by country of residence

E→T
F(1) = 5.66, \( p = .002 \), \( \eta^p = .06 \)

Covariates: age, years of usage of E and T, self ratings E and T
CLX T->E and E->T by cultural group

T->E
F(1,97) =-6.39, p = .013, η² = .065

E->T ns.
Discussion
Answers to research questions

1. Can speakers of two unrelated languages engage in congruent lexicalization (CLX)?
   **YES**, but it is the least frequent type of CS

2. Do CS patterns among groups of Turkish-English bilinguals differ, in particular in relation to CLX?
   **YES**, but need to separate out direction of CLX. CLX most frequent among UK-based Cypriots, as predicted, but:
   - Small number of extreme values in this group
   - Effect sizes small
Answers to research questions

3. Which variables predict the variance in CLX?
   - Residence (UK versus Cyprus) for E-T
   - Cultural affiliation (Cyprus versus Turkey) for T-> E.

4. Can CSFT demonstrate differences in CS patterns? Yes,
   a) high reliability
   b) possibility to investigate non-standard varieties,
   c) CSFT can reveal information about frequency of infrequent phenomena:
      Monolingual > Intersentential > INS > BFL > ALT > CLX
      But: further validation necessary (bilingual spontaneous data, more language pairs)
Where do we go from here?

- Data from Turkey-born bilinguals resident in Cyprus needed
  - Interaction between residence/cultural affiliation/place of birth
  - Interaction with language proficiency/use

- Analysis of Greek-English CSFT

- To what extent do results from CSFT correlate with task switching (executive functions) or creativity (ATTA)?

- What are the structural neural correlates of CS?

- Preliminary results suggest that frequency ratings of CLX from CSFT are positively related to brain volumes of regions involved in language switching (Korenar et al. in prep.)
References


Many thanks to CeLM, the Centre for British Research in the Levant and to all participants in the project.

More recent work on code-switching:
appendix

- Examples of different types of English-German code-switching (Hofweber et al., 2019) as found in an email discourse completion task.
Email discourse completion task - instruction

„Sie moechten einem deutsch-englischen Freund you know well vorschlagen, heute abend ins Kino zu gehn. Formulieren Sie eine kurze email, in der Sie Sprachen switchen wie im Alltagsleben. Beginnen Sie die email mit ein bisschen small talk, indem Sie erzaehlen, was Sie am Wochenende gemacht haben.“

„You would like to suggest to a **German-English friend you know well** to go to the cinema tonight. Formulate a short email in which you switch languages as you would do in everyday life. Start the email with a little small talk by telling them about what you did at the weekend.“

“Alternatively, you can copy and paste a real email you exchanged with a German-English bilingual friend into the textbox.”

**Analysis codes:**

(i) insertion  
(a) alternation  
(d) dense code-switching  
(o) other / attrition / failure to acquire  
(t) transfer / interference
Hi Name,

long time no see (a), aber es tat gut, mal wieder von dir zu hören. Wie war denn euer holiday (i) in den Bahamas? Habt ihr denn auch gesuntanned (d)? Schick doch mal ein paar snapshots (i) von euch auf der beach (d)! My old man (i) und ich haben am weekend (i) im Garten geschuftet. Der neue „Digger“ (i) kam dann sehr handy (d) und verkürzte die Arbeit by half (d). Danach, quite by chance (a), kam der Nachbar reingeschneit - wollte sich 'nen Spaten leihen, sah den Digger (i) und war blown away mit der neuen technology (d). What happened then? (a) Sure, jetzt baggert das Ding bei ihm im Garten und bei uns ist Ebbe (a)! So ein cooles Gerät will ja jeder haben, not so (a)? Lisa geht's so-so (i). Hat schon wieder Kopfschmerzen, like there's no tomorrow (a). Throbbing (a), sagt sie. Oh well (a), was soll man erwarten wenn sie jede Nacht durch partied (d)? Looks to me als ob die Jugend nie auslernt (a)! Müssen ja alle erst aus Erfahren ihre lessons lernen (d)! Und gibt sie eine helping hand (d) in der Küche? Nichts da. Mama ist ja hands on (i). War ich auch so als teenager (i)? Mensch haben unsere Eltern uns jobben lassen (d). Weißt du noch? Nichts von wegen all night out (i), und so! Ja, times are changing (a), sag ich dir! All the best (a), Bruderherz und grüß mir dein little wife (i).

Deine participant 12013