# Factors affecting the processing of compounds in the second language

Serkan Uygun & Ayşe Gürel
Yeditepe University & Boğaziçi University

# Outline

- Issue under investigation
- Introduction - What is compounding?
- Models for morphological processing
- Previous L1 and L2 studies on compound processing
- The study
- Results
- Discussion
- Conclusion

# Issue under investigation

- This study explores whether late L2 Turkish learners with L1 English process nominal compounds in the same way as native Turkish speakers do.

# What is compounding?

- Compounding is a fundamental word formation process as it offers the easiest and most effective way to create new meanings and complex words (Dressler, 2006; Libben, 2006).
- Compounding is the combination of two words, one of which modifies the meaning of the other (i.e., the head) (e.g., Bauer, 1983, 2001, 2006; Plag, 2003; Delahunty and Garvey, 2010).

# Why to study the processing of compounds?

- Compounds may consist only of two free morphemes, yet inflected and derived forms always include an affix.
- Compounds allow us to identify the role of constituents, frequency and semantic transparency in the processing of multimorphemic words (Foirentino, 2006).

# Models for morphological processing

| Model | Assumption | Prediction for Compounds |
| --- | --- | --- |
| **Full-Listing Model (Butterworth, 1983)** | Complex words are represented as whole units | Compound RT = Single word RT |
| **Decomposition Model (Taft&Forster, 1975)** | Complex words are decomposed into constituent morphemes | Compound RT > Single word RT |
| **Dual-route Models** | Frequency, family size and transparency determine which processing route applies | Full-listing for opaque, decomposition for transparent compounds |
| **The APPLE Model (Libben, 1994; 1998)** | Complex words are decomposed into constituent morphemes | Compound RT < Single word RT<br>Transparency leads to longer RT |

# L1 compound studies

- Compound studies mainly focus on the questions of:
  - whether one of the constituents has a more significant impact on the processing route;
  - whether the semantic transparency of constituents affects the parsing route

# Effects of constituents (C1and C2)

| Study | Language | Methodology | Findings | Conclusion |
|---|---|---|---|---|
| Taft&Forster, 1976 | English | Lexical decision | C1 word (footmilge) RT > C1 nonword (thernlow) RT → decomposition<br>Low frequency C1 RT > high frequency C1 RT → decomposition | C1 effects |
| Juhasz, 2006 | English | Eye movement | High frequency C1 had shorter first fixation times and gaze durations than simple words → decomposition<br>Low frequency C1 = simple words | C1 effects |
| Juhasz et al., 2003 | English | Eye movement, naming, lexical decision | Access to constituents depended on C2 frequency | C2 effects |
| Libben et al., 2003 | English | Priming | C2 opaque (staircase) RT > C2 transparent (strawberry) RT | C2 effects |
| Andrews et al., 2003 | English | Eye movement | Reliable effects of frequency for both constituents | Both C1 and C2 effects |
| Janssen et al., 2014 | English | Lexical decision | C1 and C2 frequency were found to be important | Both C1 and C2 effects |

# Semantic transparency

- Another major question: whether the semantic transparency of constituents affects the parsing route.

- Compounds are divided into four groups:
  - Transparent-transparent (TT): bedroom
  - Opaque-transparent (OT): nickname
  - Transparent-opaque (TO): shoehorn
  - Opaque-opaque (OO): deadline

- The RTs for each compound type were compared with each other to identify the effect of semantic transparency.

# Effects of semantic transparency

| Study | Language | Methodology | Findings | Conclusion |
|---|---|---|---|---|
| Libben et al., 2003 | English | Priming | Both C1 and C2 were activated regardless of semantic transparency | No effect of semantic transparency; decomposition independent of transparency |
| Jarema et al., 1999 | French | Priming | Constituents activation in all compound types | Decomposition independent of transparency |
| Sandra, 1990 | Dutch | Semantic priming | No priming effect for opaque compounds but both constituents were activated in transparent compounds | Effect of semantic transparency (dual-route) |
| Zwitserlood, 1994 | Dutch | Semantic priming | No priming effect for opaque compounds but both constituents were activated in partially and fully-transparent compounds | Effect of semantic transparency (dual-route) |
| Jarema et al., 1999 | Bulgarian | Priming | No priming effect for opaque compounds | Effect of semantic transparency (dual-route) |
| Stathis, 2014 | English | Lexical decision | Decomposition only when both C1 and C2 were transparent | Effect of semantic transparency (dual-route) |

# L2 compound processing

- Limited number of studies on L2 compound processing aim to:
  - explore how L2 learners process compound words
  - whether L2 learners differ from native speakers in terms of the route they employ in processing compounds

# How L2 learners process compounds?

| Study | Language | Methodology | Findings |
|---|---|---|---|
| Goral et al., 2008 | L1 Hebrew-L2 English | Priming | No priming effect → full-listing (for participants from Israel)<br>Priming effect → decomposition (for participants from the USA) |
| Ko, 2011 | L1 Korean-L2 English | Masked priming | No priming effect → full-listing |
| Ko et al., 2011 | L1 Korean-L2 English | Lexical decision | Compound words were decomposed into their constituents |
| Wang, 2010 | L1 Chinese-L2 English | Lexical decision | Faster RT for compounds with high frequency C2 → decomposition |
| Mayila, 2010 | L1 Chinese-L2 English | Masked priming | Decomposition for transparent, full-listing for opaque → dual-route |

# L1 processing = L2 processing?

| Study | Language | Methodology | Findings |
| --- | --- | --- | --- |
| De Cat et al., 2014; 2015 | L1 German-L2 English & L1 Spanish-L2 English | EEG recordings | Licit compounds (*coal dust*): dual-route in L1, decomposition in L2<br>Reversed compounds (*dust coal*): both groups employed decomposition |
| Li et al., 2015 | L1 Chinese-L2 English | Masked priming | L1 = L2 → decomposition<br>Both C1 and C2 were activated regardless of transparency |

# Processing studies in agglutinative languages

| Study | Language | Methodology | Findings |
|---|---|---|---|
| Kuperman et al., 2008 | Finnish | ERP | They obtained C1 frequency effect but no effect of C2 → decomposition |
| Bertram & Hyönä, 2003 | Finnish | Eye movement | Long compounds were decomposed but short compounds were stored as whole units → dual-route |
| Duñabeitia et al., 2007 | Basque | Lexical decision | They obtained C2 frequency effect High frequency C2 RT < low frequency C2 RT → decomposition |
| Vergara-Martinez et al., 2009 | Basque | ERP | They obtained C2 frequency effect High frequency C2 RT < low frequency C2 RT → decomposition |

# Processing studies of Turkish compounds

- Özer (2010) investigated three types of compounds via a morphological priming paradigm by means of a picture naming task :
  - Bare juxtaposed compounds: *akbalık* ‘dace’
  - Indefinite compounds: *dil balığı* ‘flounder’
  - Definite compounds: *gölün balığı* ‘fish of the lake’
- Picture names (e.g., *balık* ‘fish’) were morphologically related either to C1 or C2 or they were completely unrelated.
- Morphologically related compounds led to shorter naming latencies → decomposition.
- Despite not being significant, an RT advantage for C2 (head).

# The aim of the study

- The study aims to examine the representation/processing of compounds in the mental lexicon of L1-English L2-Turkish learners in comparison to Turkish native speakers.

# Compounds in Turkish

- Compounds in Turkish are formed by combining two words (Göksel, 2009).
- Turkish compounds are mostly right-headed (Yükseker, 1987; Göksel & Haznedar, 2007; Kunduracı, 2013).
- Turkish has verbal (e.g., *alay etmek* ‘to ridicule’), adjectival (e.g., *delikanlı* ‘young man’) and nominal (e.g., *büyükbaba* ‘grandfather’) compounds.

# Participants

| Groups | Gender | Mean age (Range) | Mean age of L2 exposure (Range) | Mean length of L2 exposure(Range) |
|---|---|---|---|---|
| Turkish Native Speakers (N=73) | 57F-16M | 32.37 (18-46) | At birth | From birth |
| L2 Turkish Intermediate (N=36) | 21F-15M | 40.30 (20-67) | 31.13 (17-55) | 9.08 (2-30) |
| L2 Turkish Advanced (N=35) | 24F-11M | 42.60 (21-62) | 25.14 (15-43) | 17.42 (5-40) |

# Task

- A background questionnaire
- Transparency judgment test (in a 5-point Likert scale with 86 participants)
- Turkish Placement Test of Istanbul University Language Center
- Masked priming experiment via E-prime 2.0 (Schneider, Eshman & Zuccolotto, 2002)

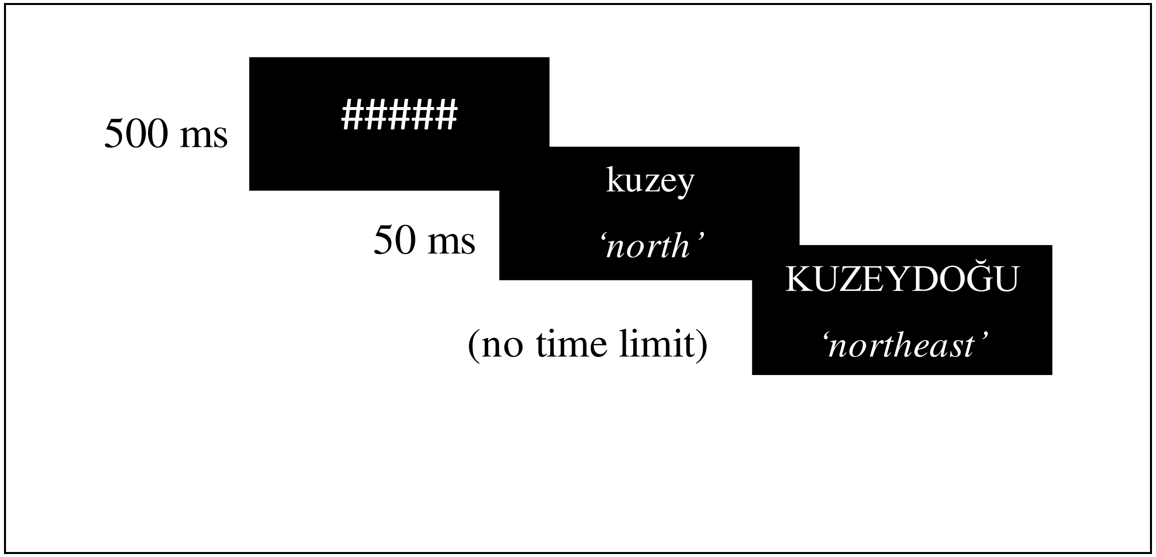
# Stimuli

| Condition | Item |
|---|---|
| Transparent-transparent (TT) | *kuzeydoğu* 'northeast' |
| Partially-opaque (PO) | *büyükelçi* 'ambassador' |
| Pseudocompound (PSC) | *fesleğen* 'basil' |
| Monomorphemic (MONO) | *kaplumbağa* 'turtle' |
| Nonword Compound | *kumardalga* |
| Nonword Monomorphemic | *ülterzatif* |

- All word items were matched on whole-word frequency, whole-word length, C1 frequency, C1 length, C2 frequency and C2 length as much as possible based on METU Corpus.

- A significant difference was only obtained between partially-opaque and pseudocompound items in terms of whole-word length ($p$=.038).

# Procedure




- The participants were asked to respond to a set of words appearing on the computer screen by pressing either a "Yes" or "No" button on the keyboard as quickly and as accurately as possible.

- The experiment automatically records the RTs and accuracy of the participants.

# Priming Conditions

| Constituent 1 (C1) | Constituent 2 (C2) | Unrelated (UR) | Target |
|---|---|---|---|
| kuzey<br>‘*north*’ | doğu<br>‘*east*’ | çanta<br>‘*bag*’ | KUZEYDOĞU<br>‘*northeast*’ |

- If RTs to the target after prime, C1 (kuzey) < UR (çanta) → C1-based decomposition
- If RTs to the target after prime, C2 (doğu) < UR (çanta) → C2-based decomposition
- If RTs to the target after prime, C1 (kuzey) = C2 (doğu) < UR (çanta) → decomposition
- If RTs to the target after prime, C1 (kuzey) = C2 (doğu) = UR (çanta) → no decomposition
- If RTs to the target after prime, C1 (kuzey) = C2 (doğu) > UR (çanta) → no decomposition

# Language test scores

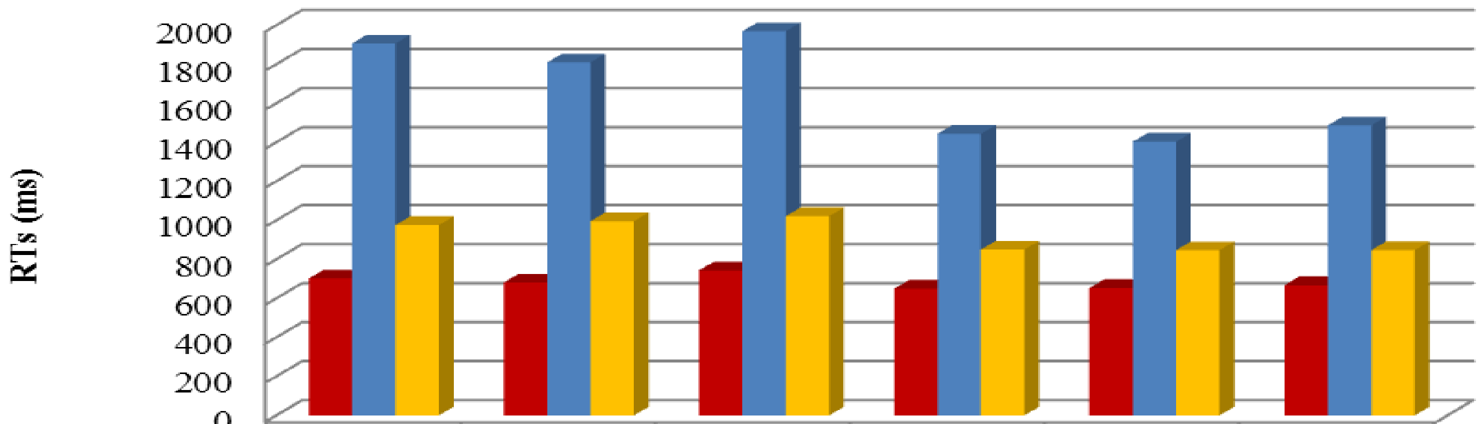
| Groups | Scores (100) | Range | Standard deviation |
|---|---|---|---|
| L2 Tr Int (N=36) | 43.50 | 33-54 | 6.98 |
| L2 Tr Adv (N=35) | 75.88 | 67.5-87 | 5.61 |

- The independent sample t-test revealed a significant difference between the Turkish test scores ($t(69)=21.447$, $p=.000$)
- Advanced level participants received significantly higher scores than the intermediate level participants.

# Data analysis

- All incorrect responses and outliers (3 SD above and below the mean) were excluded from the analysis.
- Three analyses were conducted:
  - Analysis 1: if compound words are processed differently from noncompound words
  - Analysis 2: if semantic transparency influences the processing of compounds
    - Analysis 1 and 2 also investigate which constituent has a more significant impact in processing compounds
  - Analysis 3: how pseudocompound and monomorphemic items are proccessed
- 2 (word types) x 3 (prime types) x 3 (groups) mixed-model ANOVA was conducted for each analysis.

# Compound (C) vs. Noncompound (NC)



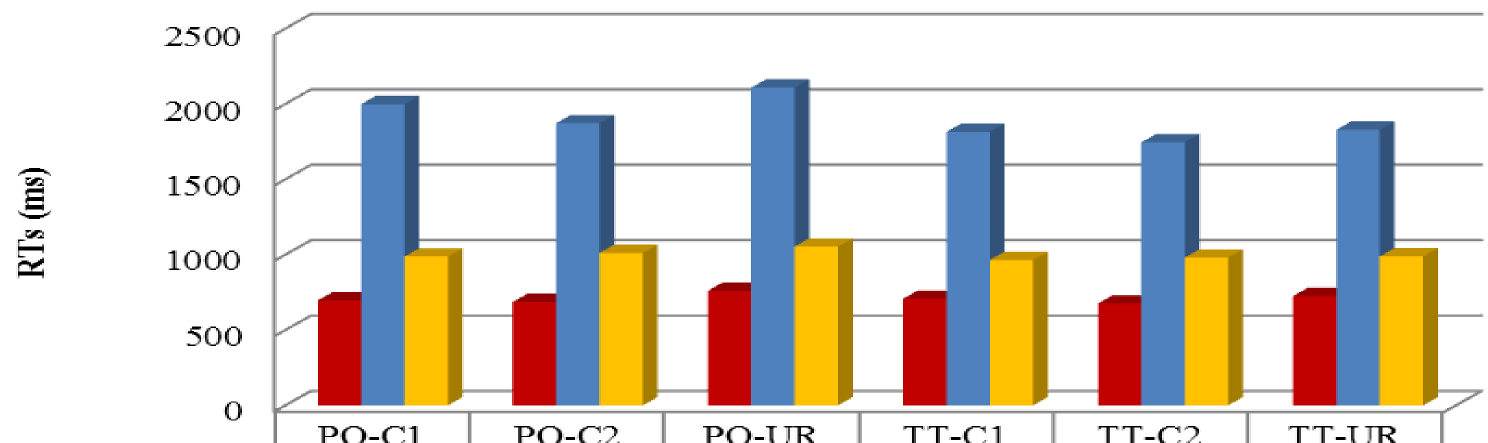

| | C-C1 | C-C2 | C-UR | NC-C1 | NC-C2 | NC-UR |
|---|---|---|---|---|---|---|
| L1 TR | 702.168 | 681.809 | 743.583 | 649.949 | 653.435 | 668.305 |
| L2 TR INT | 1906.85 | 1809.492 | 1969.394 | 1444.117 | 1403.338 | 1487.015 |
| L2 TR ADV | 975.921 | 996.106 | 1021.567 | 850.611 | 847.198 | 847.494 |

Significant differences of:
Word types ($F$=198.143; $p$=.000) → C > NC ($p$=.000)
Prime types ($F$=5.276; $p$=.006) → C2 < UR ($p$=.003)
Groups ($F$=184.691; $p$=.000) → **C:** L1 Tr < L2 Int ($p$=.000) & L2 Adv ($p$=.000), L2 Adv < L2 Int ($p$=.000); **NC:** L1 Tr < L2 Int ($p$=.000) & L2 Adv ($p$=.000), L2 Adv < L2 Int ($p$=.000)
Word types x groups ($F$=59.675; $p$=.000)

# Within-group analysis (C vs. NC)

| | L1 Turkish | L2 Turkish Int | L2 Turkish Adv |
|---|---|---|---|
| **Compound vs. Noncompound** | C > NC (*p*=.000) | C > NC (*p*=.000) | C > NC (*p*=.000) |
| **Compounds** | C1 < UR (*p*=.066)<br>C2 < UR (*p*=.002)<br>Headedness-based decomposition | No priming effects<br>→ no decomposition | No priming effects<br>→ no decomposition |
| **Noncompounds** | No priming effects<br>→ no decomposition | No priming effects<br>→ no decomposition | No priming effects<br>→ no decomposition |

# Partially-opaque (PO) vs. Transparent-transparent (TT)




| | PO-C1 | PO-C2 | PO-UR | TT-C1 | TT-C2 | TT-UR |
|---|---|---|---|---|---|---|
| L1 TR | 696.794 | 687.096 | 760.892 | 707.54 | 676.523 | 726.275 |
| L2 TR INT | 1998.981 | 1872.8 | 2109.311 | 1814.719 | 1746.184 | 1829.478 |
| L2 TR ADV | 987.34 | 1011.576 | 1054.321 | 964.502 | 980.635 | 988.812 |

Significant differences of:

Word types ($F$=10.545; $p$=.001) → PO > TT ($p$=.001)

Prime types ($F$=4.707; $p$=.01) → C2 < UR ($p$=.006)

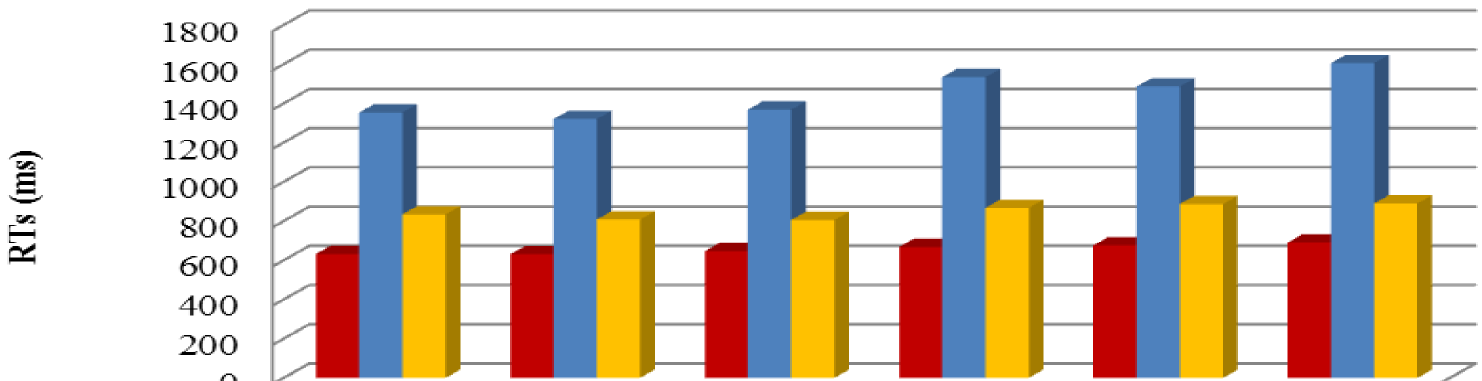
Groups ($F$=165.394; $p$=.000) → **PO:** L1 Tr < L2 Int ($p$=.000) & L2 Adv ($p$=.001), L2 Adv < L2 Int ($p$=.000); **TT:** L1 Tr < L2 Int ($p$=.000) & L2 Adv ($p$=.000), L2 Adv < L2 Int ($p$=.000)

Word types x groups ($F$=4.844; $p$=.009)

# Within-group analysis (PO vs. TT)

| | L1 Turkish | L2 Turkish Int | L2 Turkish Adv |
|---|---|---|---|
| **PO vs. TT** | No difference | No difference | No difference |
| **Partially-opaque compounds** | C1 < UR (*p*=.011)<br>C2 < UR (*p*=.006)<br>Decomposition | No priming effect<br>Full-listing | No priming effect<br>Full-listing |
| **Transparent-transparent compounds** | C2 < UR (*p*=.023)<br>Headedness-based decomposition | No priming effect<br>Full-listing | No priming effect<br>Full-listing |

# Pseudocompound (PSC) vs. Monomorphemic (MONO)




| | PSC-C1 | PSC-C2 | PSC-UR | MONO-C1 | MONO-C2 | MONO-UR |
|---|---|---|---|---|---|---|
| L1 TR | 632.153 | 630.638 | 646.118 | 667.745 | 676.232 | 690.492 |
| L2 TR INT | 1354.103 | 1320.414 | 1368.386 | 1534.131 | 1486.262 | 1605.643 |
| L2 TR ADV | 834.126 | 807.659 | 805.09 | 867.0964 | 886.738 | 889.899 |

Significant differences of:

Word types ($F$=69.975; $p$=.000) → PSC < MONO ($p$=.000)

Groups ($F$=184.050; $p$=.000) → L1 Tr < L2 Int ($p$=.000) & L2 Adv ($p$=.000), L2 Adv < L2 Int ($p$=.000)

Word types x groups ($F$=14.684; $p$=.000)

# Within-group analysis (PSC vs. MONO)

| | L1 Turkish | L2 Turkish Int | L2 Turkish Adv |
|---|---|---|---|
| **PSC vs. MONO** | PSC < MONO ($p$=.000) | PSC < MONO ($p$=.000) | PSC < MONO ($p$=.000) |
| **Pseudocompound** | No priming effect Full-listing | No priming effect Full-listing | No priming effect Full-listing |
| **Monomorphemic** | No priming effect Full-listing | No priming effect Full-listing | No priming effect Full-listing |

# Discussion

| L1 Turkish | L2 Turkish |
|---|---|

**L1 Turkish**

- Compounds are accessed in a decomposed fashion.
- Semantic transparency:
  - Both constituents are activated in PO
  - Headedness-based decomposition for TT
- Noncompounds are stored as unanalyzed whole units.

**L2 Turkish**

**L2 Tr Int**

- No priming effect → full-listing for all items

**L2 Tr Adv**

- No priming effect → full-listing for all items

- Noncompounds are stored as unanalyzed whole units.

# Conclusion

| L1 Turkish | L2 Turkish |
|---|---|

**L1 Turkish**

- Decomposition
  - Longer RTs for compounds
  - Opaqueness → activation of both constituents

**L2 Turkish**

- They are not native-like in compound processing.
- Factors that may influence native-like processing:
  - Age of L2 exposure
  - Familiarity with the constituents
  - Formal instruction

# Recommendations for further research

- Using different experimental techniques such as eye tracking
- Investigating the processing of fully-opaque compounds
- Investigating L2 Turkish participants who have similar lengths of L2 Turkish exposure
- Measuring L2 Turkish participants' proficiency level with a standardized proficiency test

# Acknowledgment

This study has been supported by a research grant given to Ayşe Gürel by the Scientific and Technological Research Council of Turkey (TÜBİTAK-1001) (research grant no:112K183).