

CONTEXTUAL DIVERSITY AS A LEXICAL ORGANIZER OF MULTIWORD EXPRESSIONS

Marco S. G. Senaldi, Debra Titone and Brendan T. Johns

Department of Psychology, McGill University

CSBBCS 2022 – 20 July 2022 – Halifax, NS



McGill

EXPRESSIONS FROM EVERYDAY LANGUAGE



crossing our
fingers...

...to not **get stuck** in
a **traffic jam**...



...while meeting a **close**
friend to **spill some tea**

MULTIWORD EXPRESSIONS

- Increased availability of corpora showed most of our linguistic production is **routinized** and **prefabricated**
 - Christiansen & Arnon 2017, Jackendoff 1997, VanLancker-Sidtis & Rallon 2004
- **Multiword expressions, formulaic units or phraseological units**
 - Arnon & Snider 2010, Siyanova-Chanturia et al. 2011, Wray 1999
 - collocations (*torrential rain*)
irreversible binomials (*black and white*)
phrasal verbs (*catch up*)
idioms (*break the ice*)
etc.

IDIOMS



on cloud nine

on

cloud

nine

break the ice

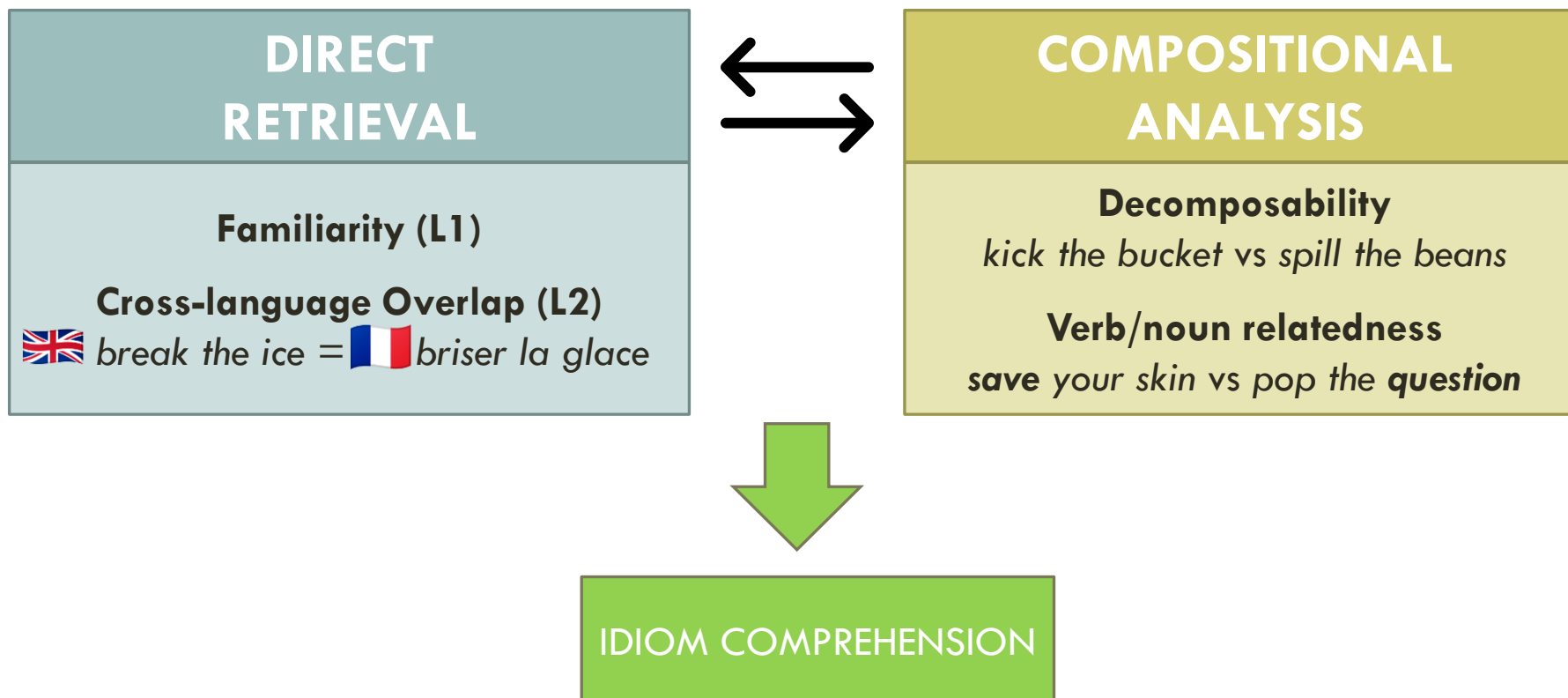
break

the

ice

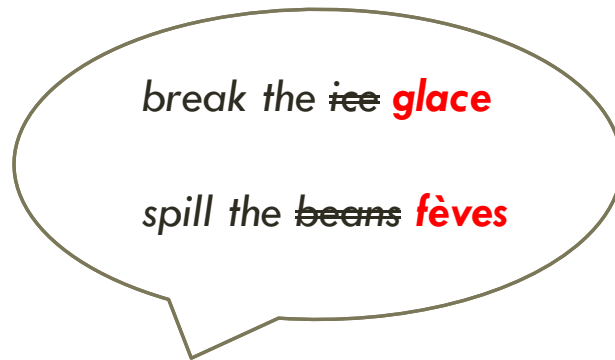
IDIOM PROCESSING: A HYBRID MATTER?

Libben & Titone 2008, *Mem Cogn*; Titone & Libben 2014, *Ment Lex*; Titone et al. 2019, *CJEP*



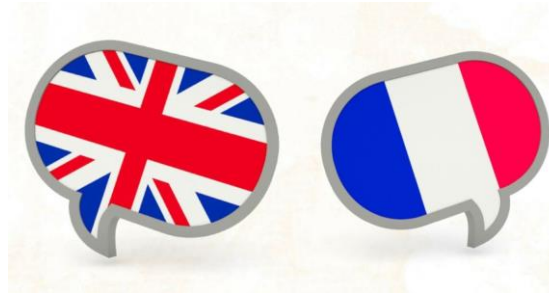
1

How are modified idioms processed?



2

How are idioms processed in an L2?



3

What do corpus measures of contextual diversity reveal about direct retrieval?



3 OPEN QUESTIONS

1

**How are modified idioms
processed?**

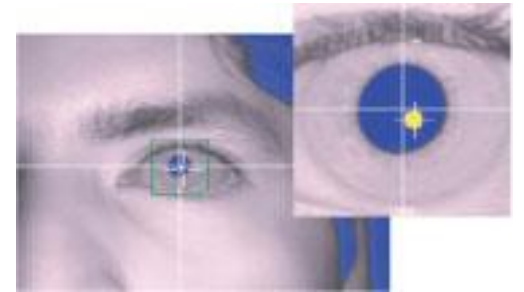
break the ~~ice~~ **glace**

spill the ~~beans~~ **fèves**

3 OPEN QUESTIONS

1. PROCESSING LANGUAGE-SWITCHED IDIOMS

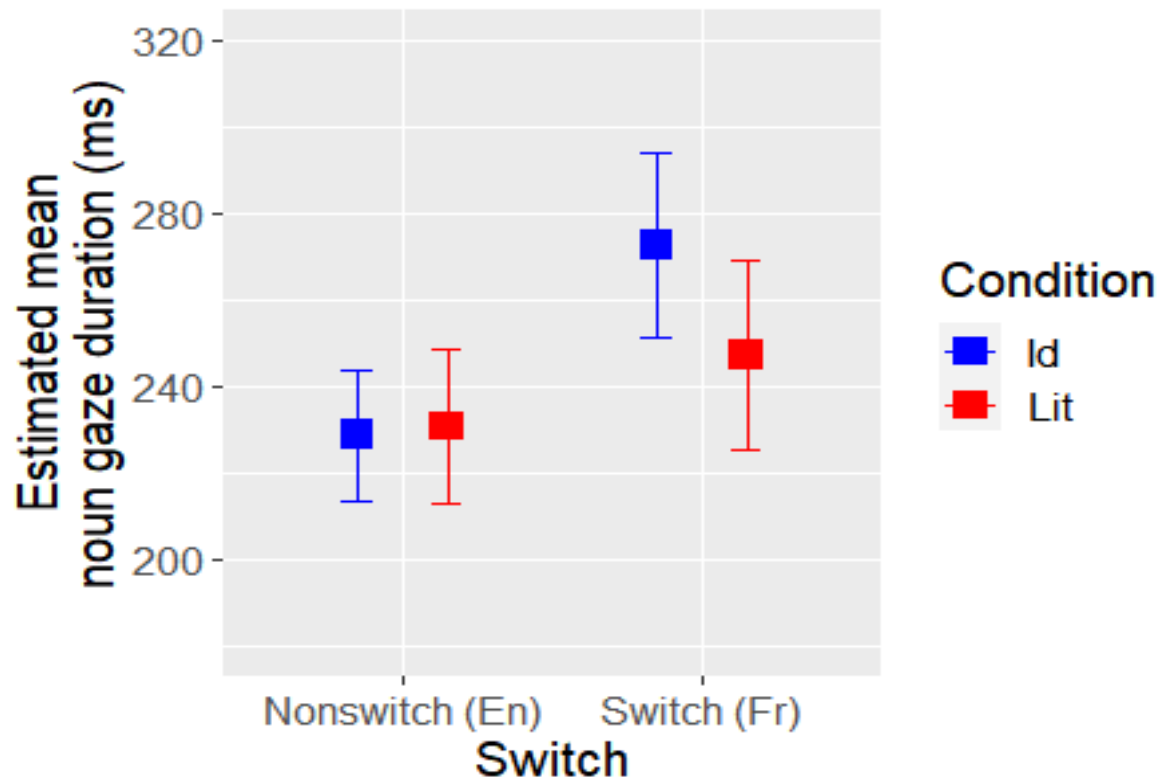
- Common feature of multilingual language processing
- **Language-switching idiom final nouns** disrupts the common idiomatic form while minimally altering the meaning and syntax of the original constituents
 - *scratch your **head*** → *scratch your **tête***
- **Eye-tracking reading experiment**
 - **41 English-French bilinguals** from Montreal **reading in their L1**
 - 42 English verb-determiner-noun idioms
 - English idioms and controls with final nouns intact or switched to French



1. STIMULI

Condition	Stimulus
NONSWITCH	
Id-Id-En	Niles <u>scratched his head</u> when he could not determine where his dog was hiding.
Id-Lit-En	Niles <u>scratched his head</u> when he returned from camp covered in mosquito bites.
Lit-Lit-En	Niles <u>examined his head</u> when he returned from camp covered in mosquito bites.
SWITCH	
Id-Id-Fr	Niles <u>scratched his tête</u> when he could not determine where his dog was hiding.
Id-Lit-Fr	Niles <u>scratched his tête</u> when he returned from camp covered in mosquito bites.
Lit-Lit-Fr	Niles <u>examined his tête</u> when he returned from camp covered in mosquito bites.

1. CONDITION X SWITCH (NOUN GD)



Switch > nonswitch

Switched idioms > switched literals

Nonswitch

Niles scratched his head ...

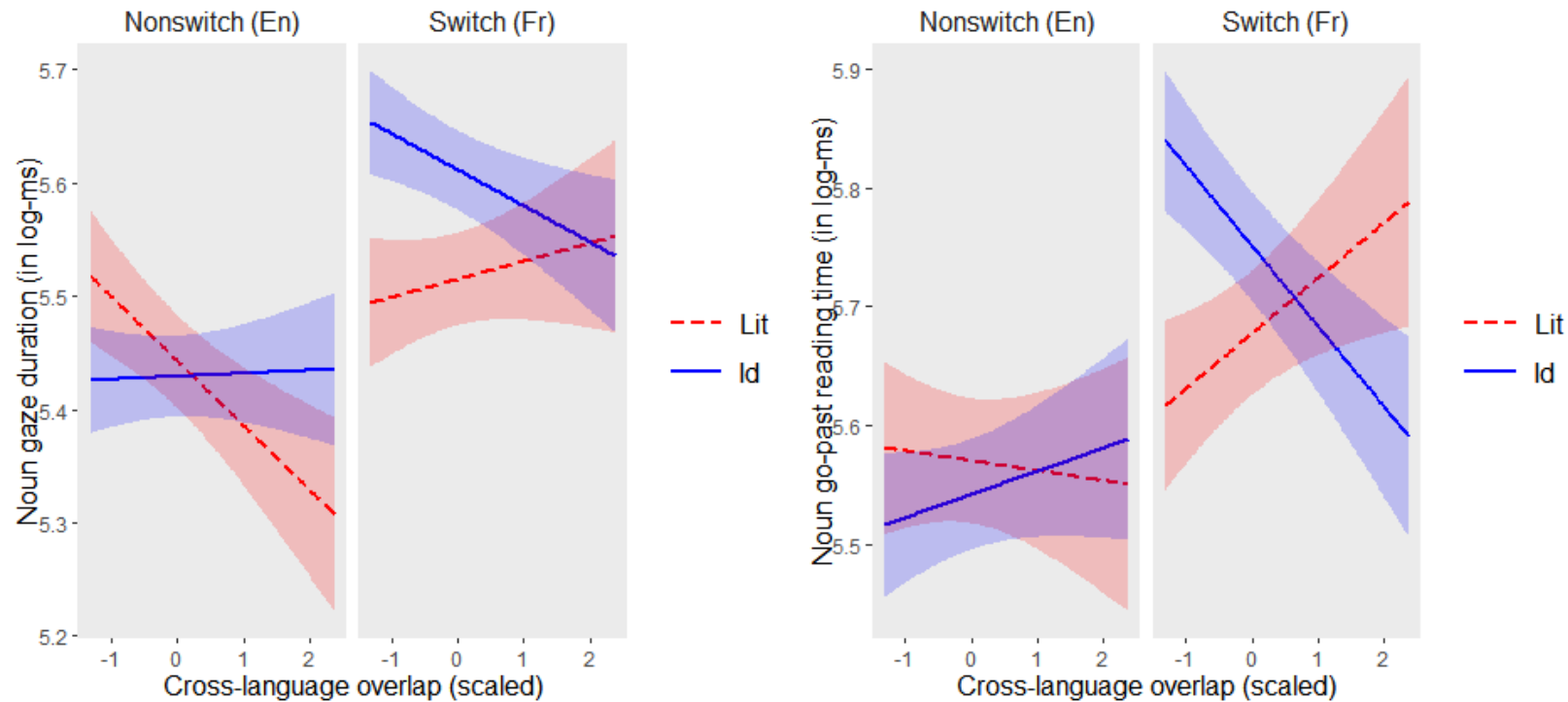
Niles examined his head ...

Switch

Niles scratched his tête ...

Niles examined his tête ...

1. OVERLAP, CONDITION AND SWITCH (NOUN GD, NOUN GO-PAST)



Switched idioms facilitated by **overlap** (*break the ice – briser la glace*)

Nonswitch

Niles *scratched his head* ...

Niles *examined his head* ...

Switch

Niles *scratched his tête* ...

Niles *examined his tête* ...

1. DISCUSSION

Direct retrieval paramount to L1 idiom recognition

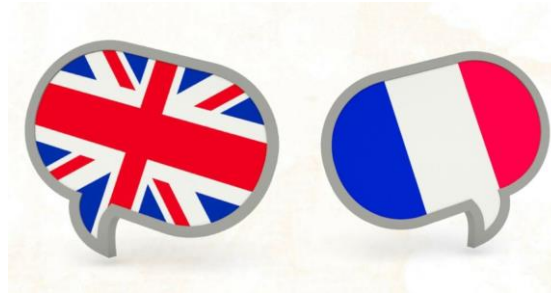
Also adopted as a repair strategy when canonical form is not directly available

Bilinguals can retrieve congruent idiomatic forms from their non-target language

Bilingual lexicon integrated beyond the single-word level?

2

**How are idioms processed
in an L2?**



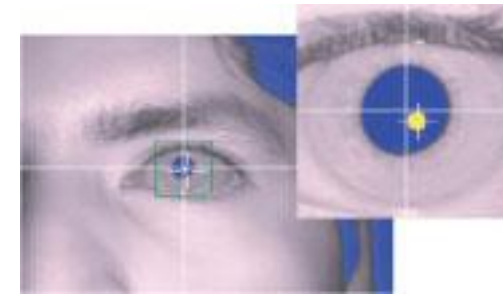
3 OPEN QUESTIONS

2. PROCESSING IDIOMS IN THE L2

- **General disadvantage** for L2 speakers in learning, producing and comprehending idioms
 - increased priming for literal than idiomatic targets (Cieślicka 2006; Beck & Weber 2016)
 - literal meaning of ambiguous idioms read faster (Siyanova-Chanturia et al. 2011)
- In line with **usage-based models**
 - Bybee 2006; Goldberg 2006; Tomasello 2003
 - ambient exposure of L2 users to idiomatic language does not suffice to form entrenched representations that can be retrieved directly
- How is the interplay of direct retrieval and compositional parsing affected in L2 processing?

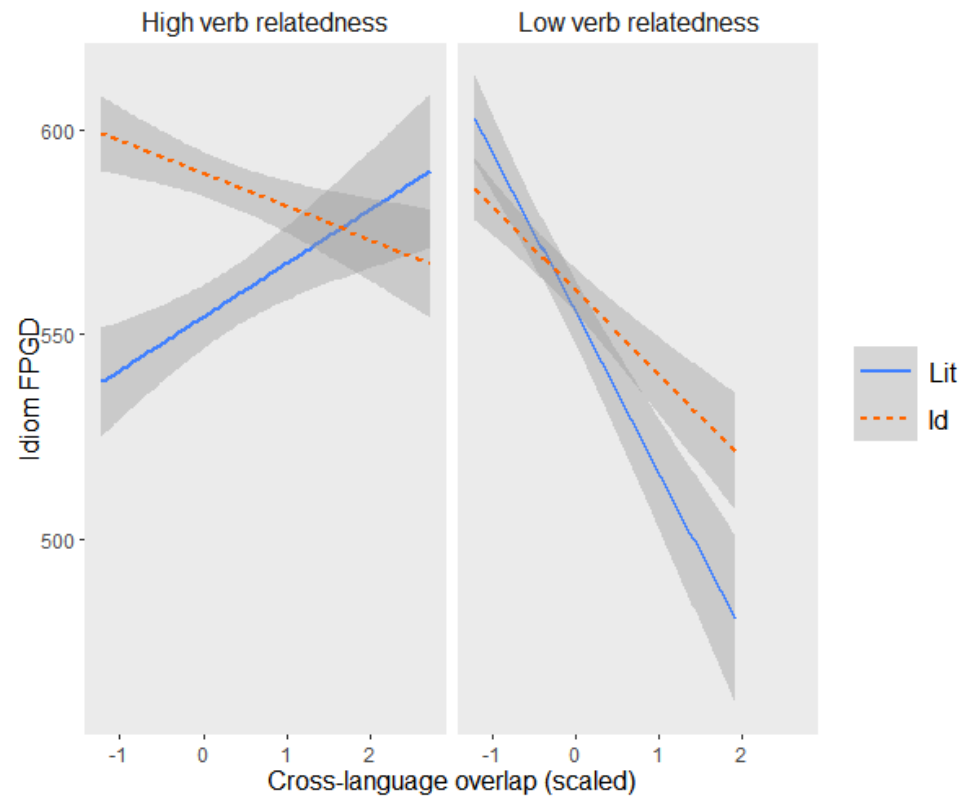
2. PROCESSING IDIOMS IN THE L2

- Eye-tracking reading experiment
- **37 French-English bilinguals** from Montreal
- 60 English verb-determiner-noun idioms



Condition	Stimulus
Id-Id	Niles <u>scratched his head</u> when he could not determine where his dog was hiding.
Id-Lit	Niles <u>scratched his head</u> when he returned from camp covered in mosquito bites.
Lit-Lit	Niles <u>examined his head</u> when he returned from camp covered in mosquito bites.

2. VERB RELATEDNESS (IDIOM FPGD)

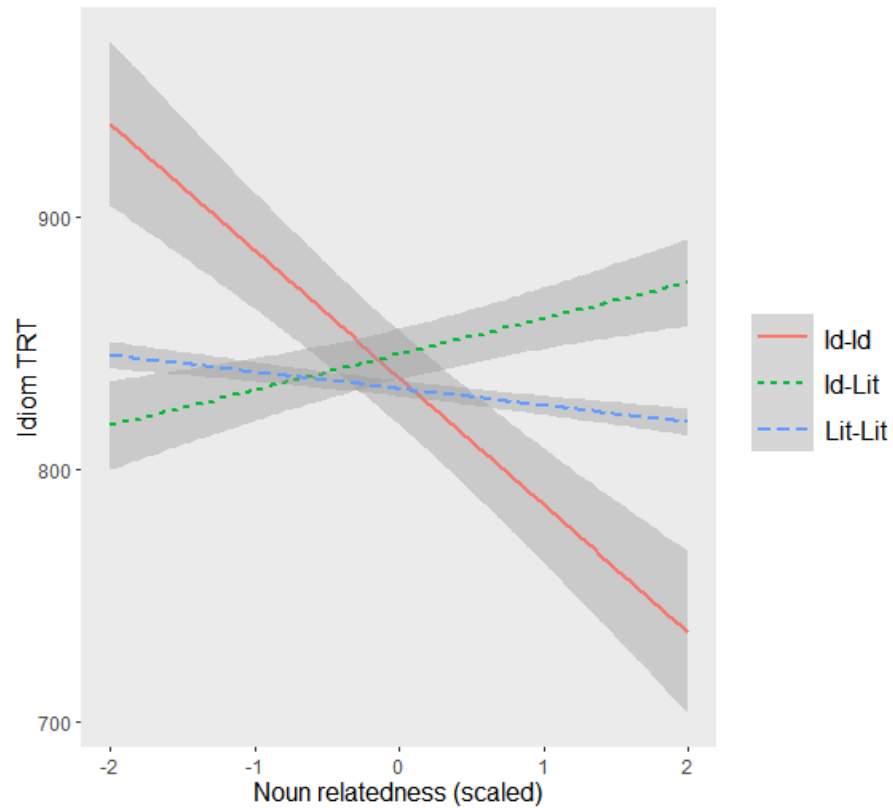


↓ **overlap** + ↑ **verb relatedness** = **idiomatic priming**

break the ice save your skin
 =
briser la glace

Condition	Stimulus
Id-Id	Niles <u>scratched his head</u> when he could not determine where his dog was hiding.
Id-Lit	Niles <u>scratched his head</u> when he returned from camp covered in mosquito bites.
Lit-Lit	Niles <u>examined his head</u> when he returned from camp covered in mosquito bites.

2. NOUN RELATEDNESS (IDIOM TRT)

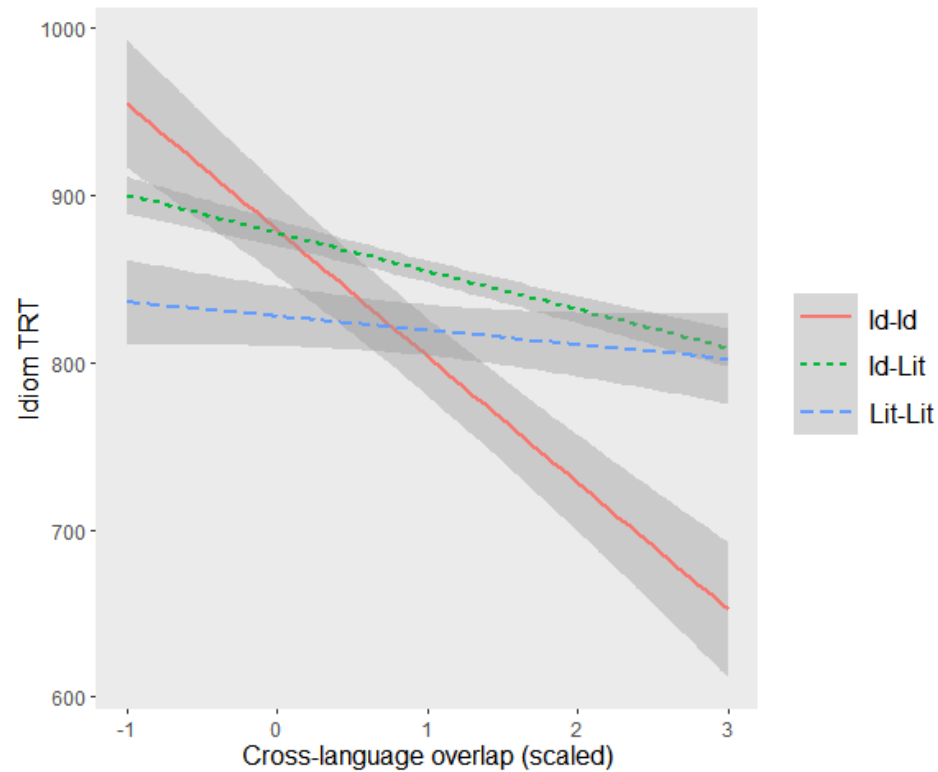


↑ Noun relatedness = late L2 idiomatic facilitation

pop the question

Condition	Stimulus
Id-Id	Niles <u>scratched his head</u> when he could not determine where his dog was hiding.
Id-Lit	Niles <u>scratched his head</u> when he returned from camp covered in mosquito bites.
Lit-Lit	Niles <u>examined his head</u> when he returned from camp covered in mosquito bites.

2. CROSS-LANGUAGE OVERLAP (IDIOM TRT)



↑ **Overlap** = **late L2 idiomatic facilitation**

break the ice
=
briser la glace

Condition	Stimulus
Id-Id	Niles <u>scratched his head</u> when he could not determine where his dog was hiding.
Id-Lit	Niles <u>scratched his head</u> when he returned from camp covered in mosquito bites.
Lit-Lit	Niles <u>examined his head</u> when he returned from camp covered in mosquito bites.

2. DISCUSSION

L2 idiom processing is mostly compositional

Early-stage figurative priming guided by prospective decomposability

Late-stage processing facilitated by retrospective decomposability

≠ L1 idiom processing, direct retrieval happens later and from the non-target language

3

What do corpus measures of contextual diversity reveal about direct retrieval?



3 OPEN QUESTIONS

3. CONTEXTUAL AND SEMANTIC DIVERSITY AS LEXICAL ORGANIZERS

- **Classical frequency-based** accounts of the lexicon (Broadbent, 1967)
 - **Principle of repetition**

There are only eight living species of **bear**, although the **bear** lives in a widespread habitat.

...

The **bear** is a solitary animal, with the exception of courting or a mother **bear** with her young.

...

The **bear** is an expert runner, climber, and swimmer.

...

People have hunted **bear** since prehistoric time, and **bear's** play a prominent role in many cultures mythology.

3. CONTEXTUAL AND SEMANTIC DIVERSITY AS LEXICAL ORGANIZERS

- Better fit of **contextual diversity (CD)** to **single-word lexical behavior data**
 - Word repetition in context is ignored (Adelman, Brown & Quesada, 2006)
 - Principle of **likely need** (Anderson & Milson, 1989)
- **Semantic Distinctiveness Model** (Jones, Johns & Recchia 2012)
 - Accounts for semantic redundancy in contexts
 - **Expectancy-congruency**
 - c = context vector
 M = memory matrix
 $SD = e^{-\lambda * S(Mi, c)}$

There are only eight living species of **bear**, although the bear lives in a widespread habitat.

...

The **bear** is a solitary animal, with the exception of courting or a mother bear with her young.

...

The **bear** is an expert runner, climber, and swimmer.

...

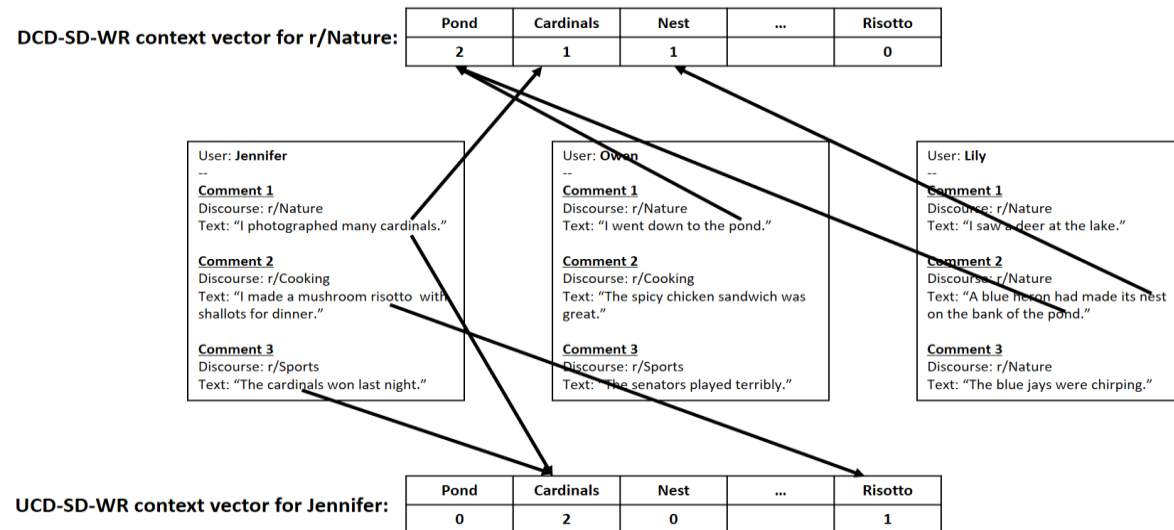
People have hunted **bear** since prehistoric time, and bear's play a prominent role in many cultures mythology.

3. SOCIALLY-BASED CONTEXTUAL DIVERSITY

- **Communication patterns of individuals across discourses** (Johns 2021)
 - 55-billion-word corpus of Reddit comments
 - **benchmark fit to single-word measures of lexical access** (lexical decision, naming, etc.)
- **Count models**
 - Discourse Contextual Diversity (DCD) = number of subreddits a word appears in
 - User Contextual Diversity (UCD) = number of users that use a word

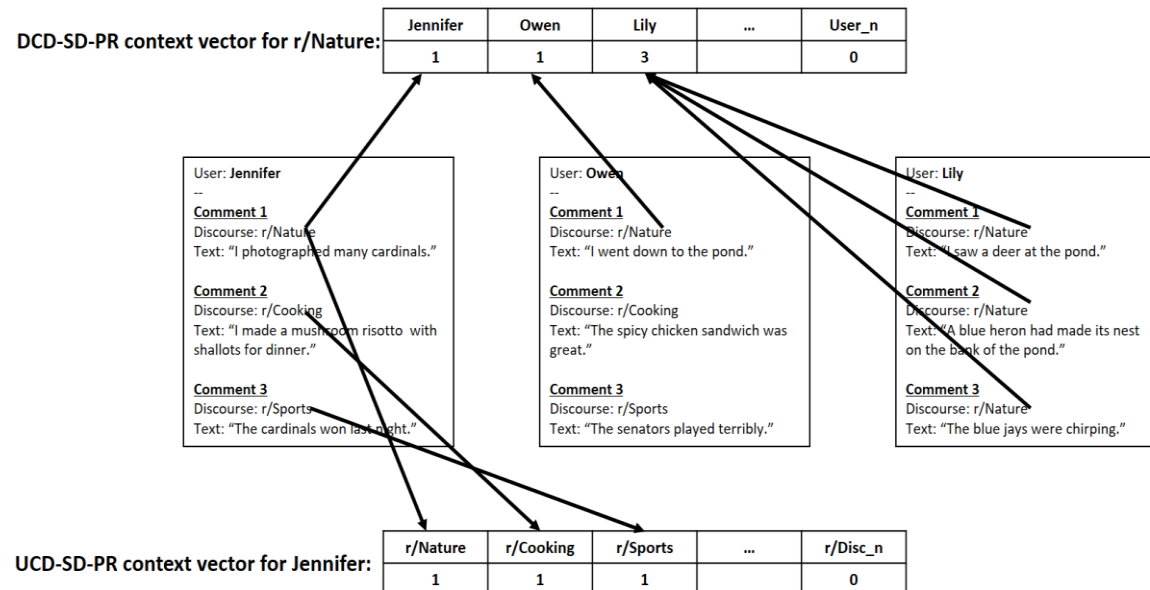
3. SOCIALLY-BASED SEMANTIC DIVERSITY: WORD REPRESENTATION MODELS

- **Discourse Semantic Diversity (DCD.SD.WR)** = to what extent a word occurs across subreddits containing diverse words
- **User Semantic Diversity (UCD.SD.WR)** = to what extent a word is used by users who have a diverse vocabulary



3. SOCIALLY-BASED SEMANTIC DIVERSITY: POPULATION REPRESENTATION MODELS

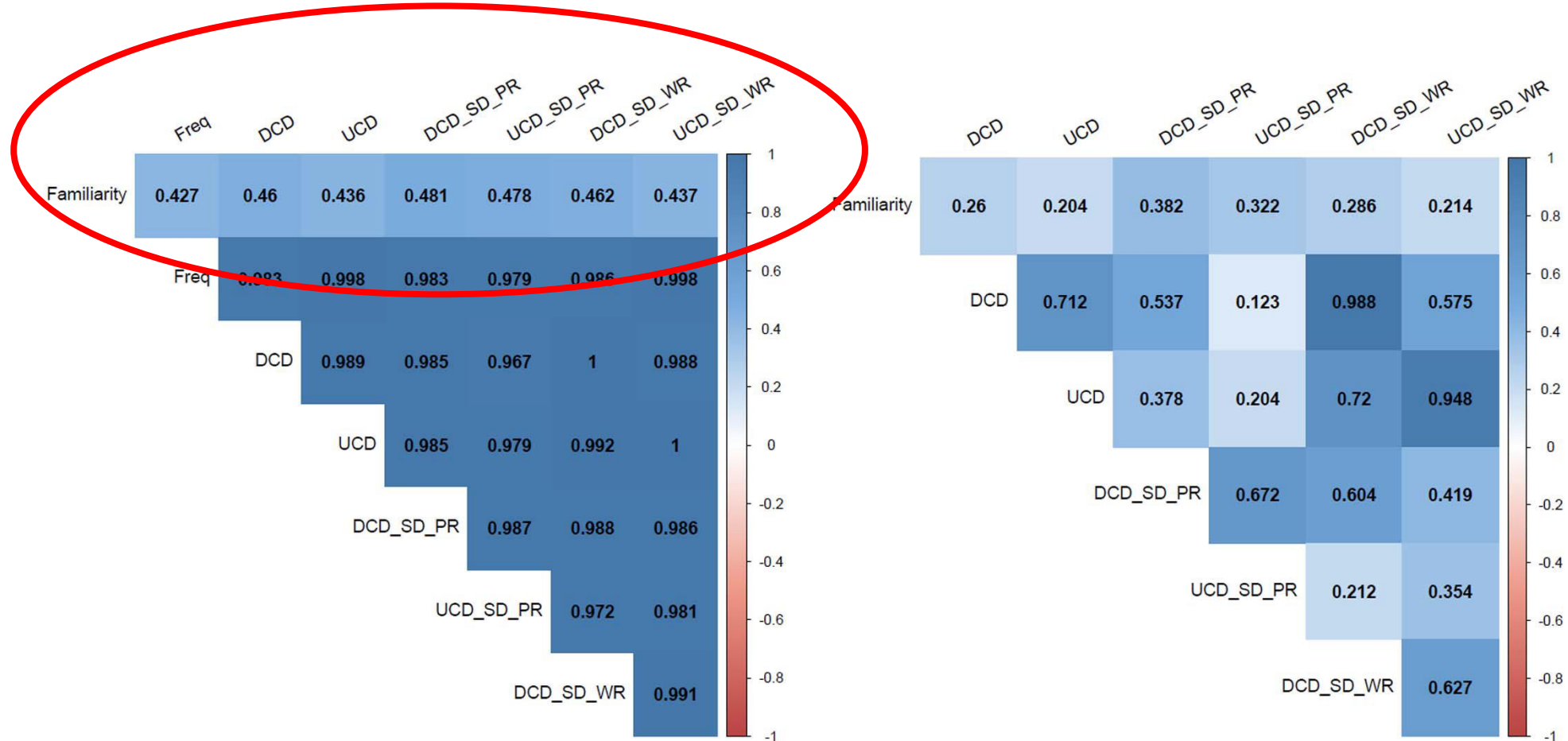
- **Discourse Semantic Diversity (DCD.SD.PR)** = to what extent a word occurs across many subreddits commented on by a diverse pool of users
- **User Semantic Diversity (UCD.SD.PR)** = to what extent a word is used by many users who tend to comment on a diverse range of subreddits



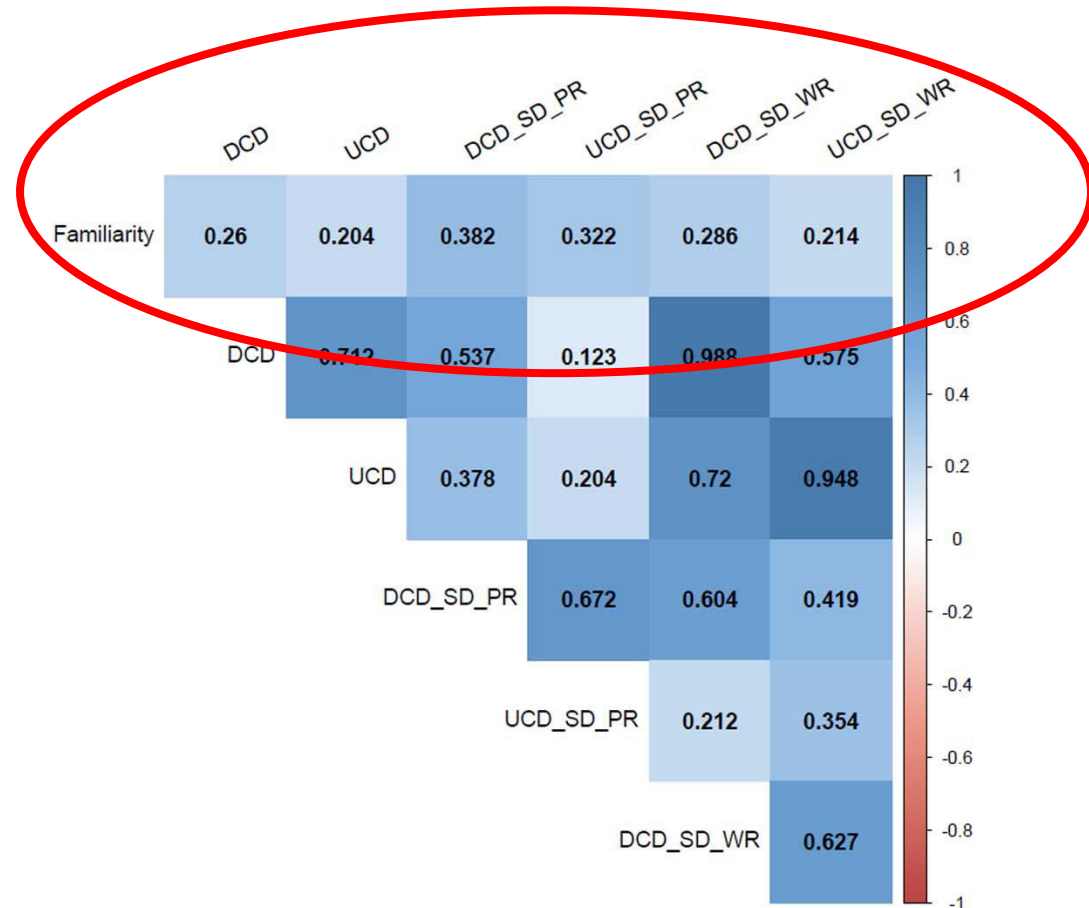
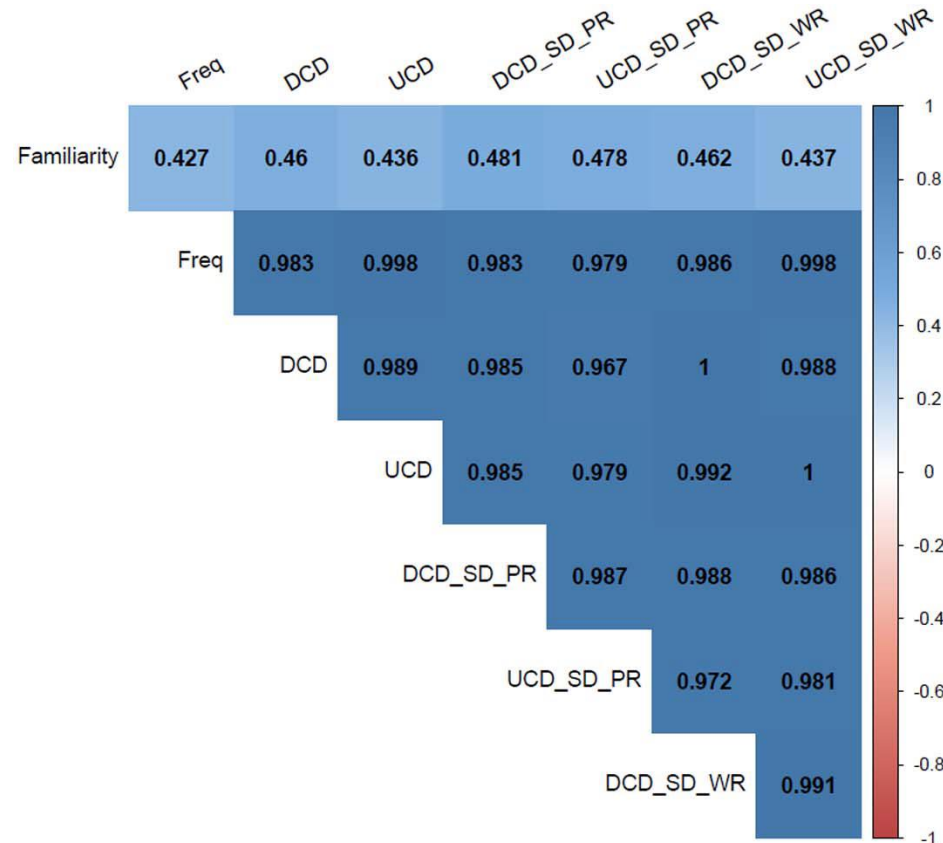
3. DO DIVERSITY MEASURES BETTER ACCOUNT FOR IDIOMS' LEXICAL ORGANIZATION?

- 210 verb-determiner-noun English idioms (Libben & Titone 2008)
- Idioms rated on **familiarity**
- Contextual, semantic and socially-based diversity computed for the idioms on the Reddit corpus
- Corpus-based semantic diversity measures compared against familiarity ratings
 - Simple and partial correlations
 - Hierarchical regressions

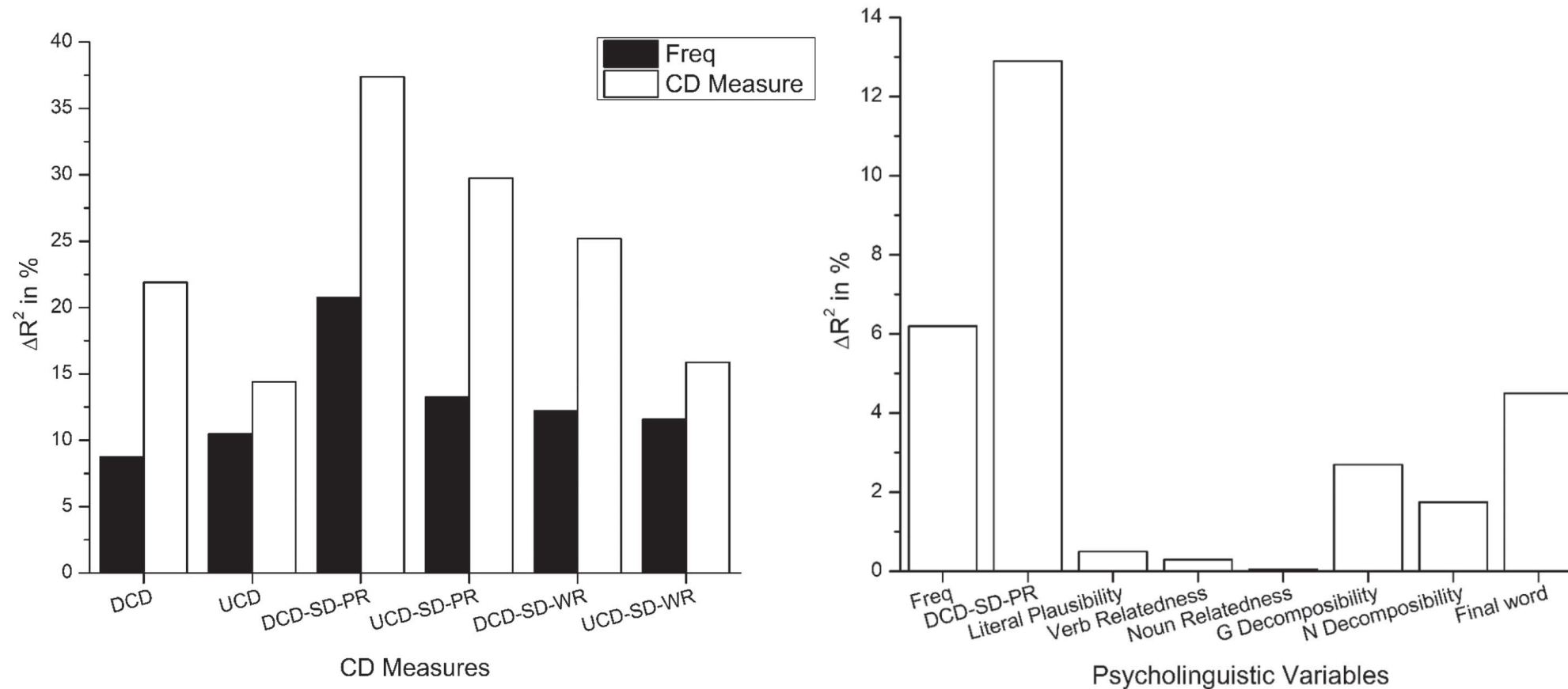
3. CORRELATIONS OF CONTEXTUAL DIVERSITY WITH IDIOMS' FAMILIARITY RATINGS



3. CORRELATIONS OF CONTEXTUAL DIVERSITY WITH IDIOMS' FAMILIARITY RATINGS



3. UNIQUE VARIANCE ACCOUNTED FOR BY DIVERSITY MEASURES



3. DISCUSSION

Variable modulating idioms' direct retrieval better predicted by socially-based contextual diversity

Idioms show lexical organization dynamics similar to single words

More familiar idioms are used across different discourses, which are commented on by an unpredictable pool of users

Phrasal frequency still plays a role, as compared to single words

1

How are modified idioms processed?

When idioms' canonical form is disrupted, readers can still rely on direct retrieval as a repair strategy

2

How are idioms processed in an L2?

L2 readers rely more heavily on compositional analysis, with direct retrieval playing a secondary role

3

What do corpus measures of contextual diversity reveal about direct retrieval?

Computational data suggests ease of direct retrieval depends on how likely an idiom is needed in a context

WRAPPING UP

WHAT LIES AHEAD

How are idioms processed after
a biasing context?

Can hybrid models account for
idiom production too?

Can contextual diversity account
for idioms' **on-line processing**?

Bruce, **who was typically shy**, broke the ice
by quickly introducing himself

Bruce, **who was an outdoorsman**, broke the
ice by driving his snowmobile on the lake.



THANK YOU!



Marco S. G. Senaldi
marco.senaldi@mcgill.ca



Debra Titone
debra.titone@mcgill.ca



Brendan T. Johns
brendan.johns@mcgill.ca

