The impact of information structure on the emergence of differential object marking: an experimental study

Shira Tal, Kenny Smith, Jennifer Culbertson,
Eitan Grossman & Inbal Arnon

In this talk

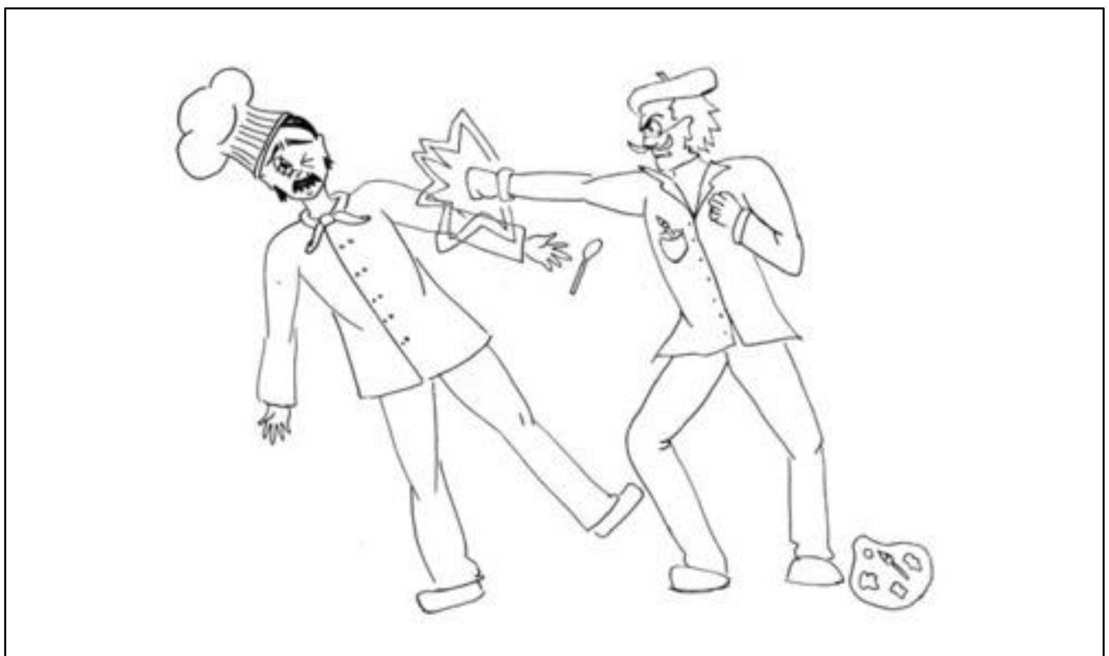
- Emergence of differential object marking (DOM) systems in the lab
- Hypothesis 1: DOM driven by disambiguation
- Hypothesis 2: DOM driven by information structure
- Some evidence that H1 drives DOM in the lab…but H2 is how most DOM arises in natural language!

Differential object marking

- Differential object marking languages mark only certain types of direct objects
  - Over 300 languages (Bossong, 1985; e.g., Turkish, Persian, Hebrew, Malayalam, Spanish)
  - E.g, definite objects in Hebrew

What is the function of DOM?

1. Distinguishing function: used in order to distinguish the agent from the patient (Comrie, 1989; Dixon, 1979; Dixon 1994)
   - Should be used more when there is ambiguity

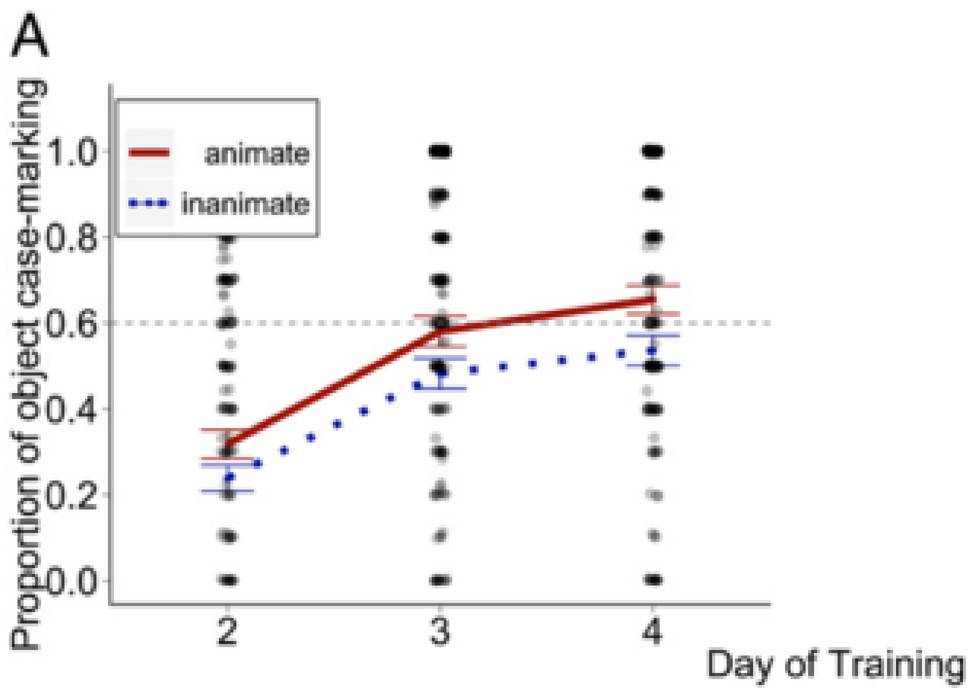


Fedzechkina et al. 2012

- Exposed learners to an artificial language
  - Flexible word order (60% SOV, 40% OSV)
  - Optional case marking on objects (60%)
    - Unconditioned: both orders, animate/inanimate (subjects always animate)
- Prediction:
  - Learners will introduce animacy-contingent case marking to reduce ambiguity:
    - Increase marking for animate objects
    - Decrease marking for inanimate objects

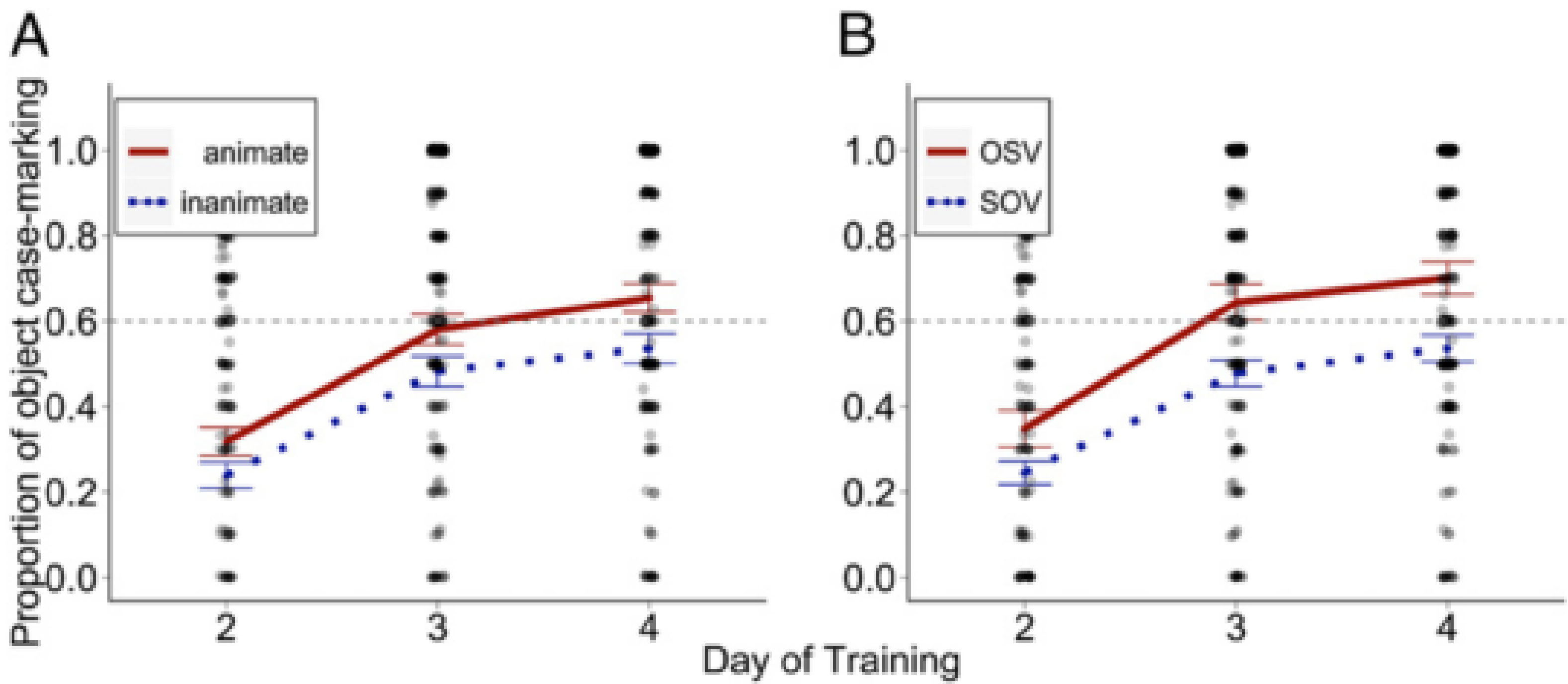
Fedzechkina, Jaeger & Newport, 2012

Learners changed their input to make it less ambiguous



Increased case marking on animate objects and reduced it on inanimate objects

Learners changed their input to make it less ambiguous



Increased case marking for OSV productions

Challenges for the distinguishing function

- The pressure to disambiguate might be less pronounced in actual languages (Aissen 2003; DuBois, 1987, Levshina, in prep; Malchukov, 2008; Seržant, 2019)

- The distinguishing function rarely drives the historical development of DOM (Cristofaro , 2013; De Hoop & Malchukov, 2008; Seržant, 2019)

What is the function of DOM?

1. Distinguishing function: used in order to distinguish the agent from the patient (Comrie, 1989; Dixon, 1979; Dixon 1994)

2. Indexing function: highlights atypical alignments (Hopper & Thompson, 1980; Næss, 2004; Næss 2007)
   - The man pushed the door
   - The door pushed the man

- A case of **Markedness/Frequency asymmetries**: mark less expected forms with more linguistic material (Haspelmath 2008; forthcoming; Levshina & Witzlack-Makarevich, in prep)

Marking of atypical alignments

- Objects could be atypical in many ways
  - Animacy: most objects are **inanimate**
    - Animate objects are atypical
  - Information structure: most objects are **new**
    - Subjects (but not objects) tend to be given

(Bossong, 1991; Croft, 2003; DuBois, 1987; 2003 ; Givón, 1976; Silverstein, 1976)

Information structure and DOM

- Object marking in different languages is impacted by information structure (e.g., Catalan, Persian) (Iemmolo, 2010; 2011; 2013; von Heusinger & Kaiser, 2007, 2011; Escandell-Vidal, 2009; Dalrymple & Nikolaeva, 2011)
  - **Given/topical objects are marked more**
  - Primary driving force for DOM cross-linguistically (Iemmolo, 2011;2013)

- Suggested as a diachronic source for DOM (Rohlfs, 1971; Iemmolo, 2010)

Our research question

Does DOM emerge from information structure alternations **in the lab**?

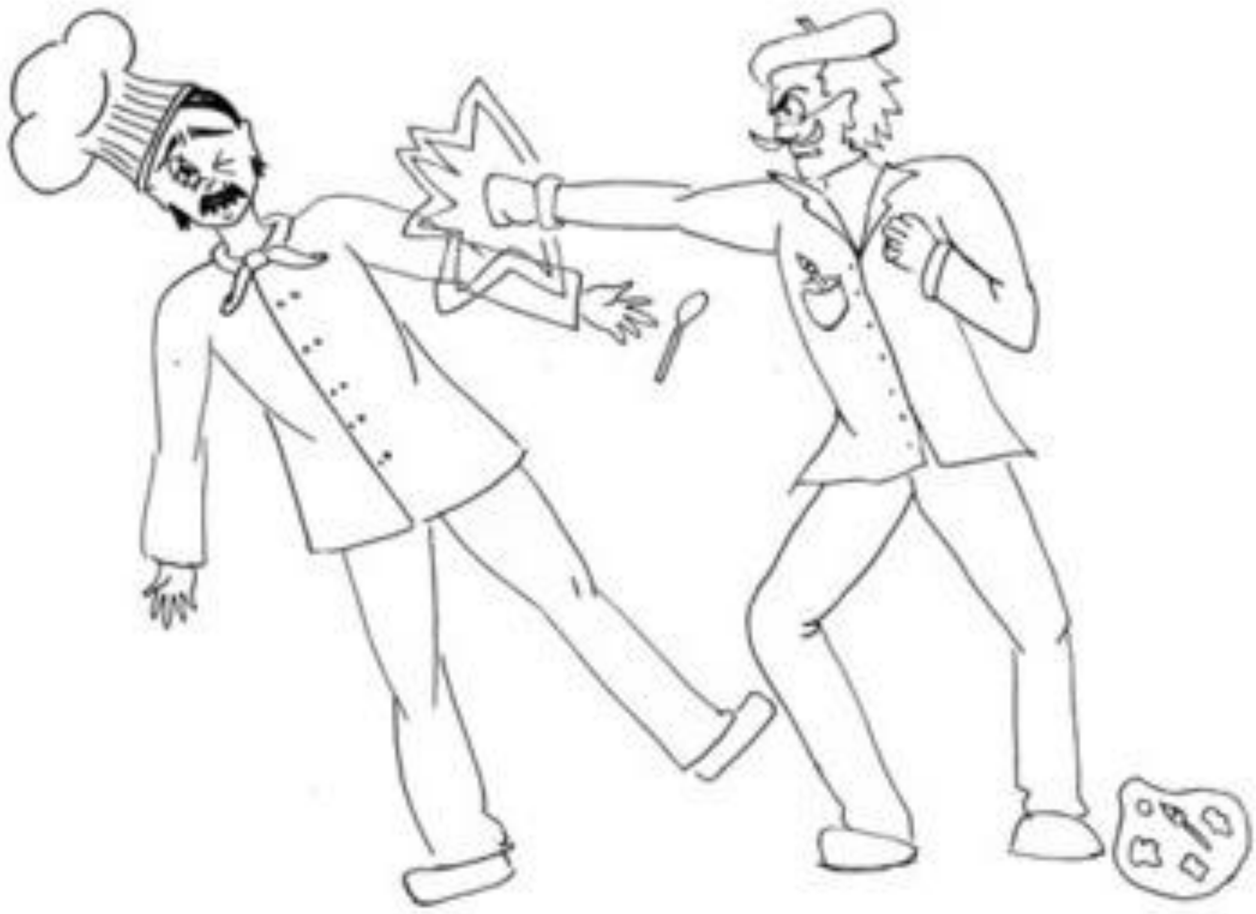
⇒ Will speakers condition case marking on givenness of objects?

Experimental Design

- Taught 43 participants an artificial language over 4 consecutive days (à la Fedzechkina et al.). Run on MTurk.

- 8 characters (all animates), 4 verbs, one optional (50%) case marking (only on objects), two word orders: OSV, SOV (50% each)

- Before each sentence, one of the characters (subject/object, 50%) is introduced (thereby becoming ‘given’)

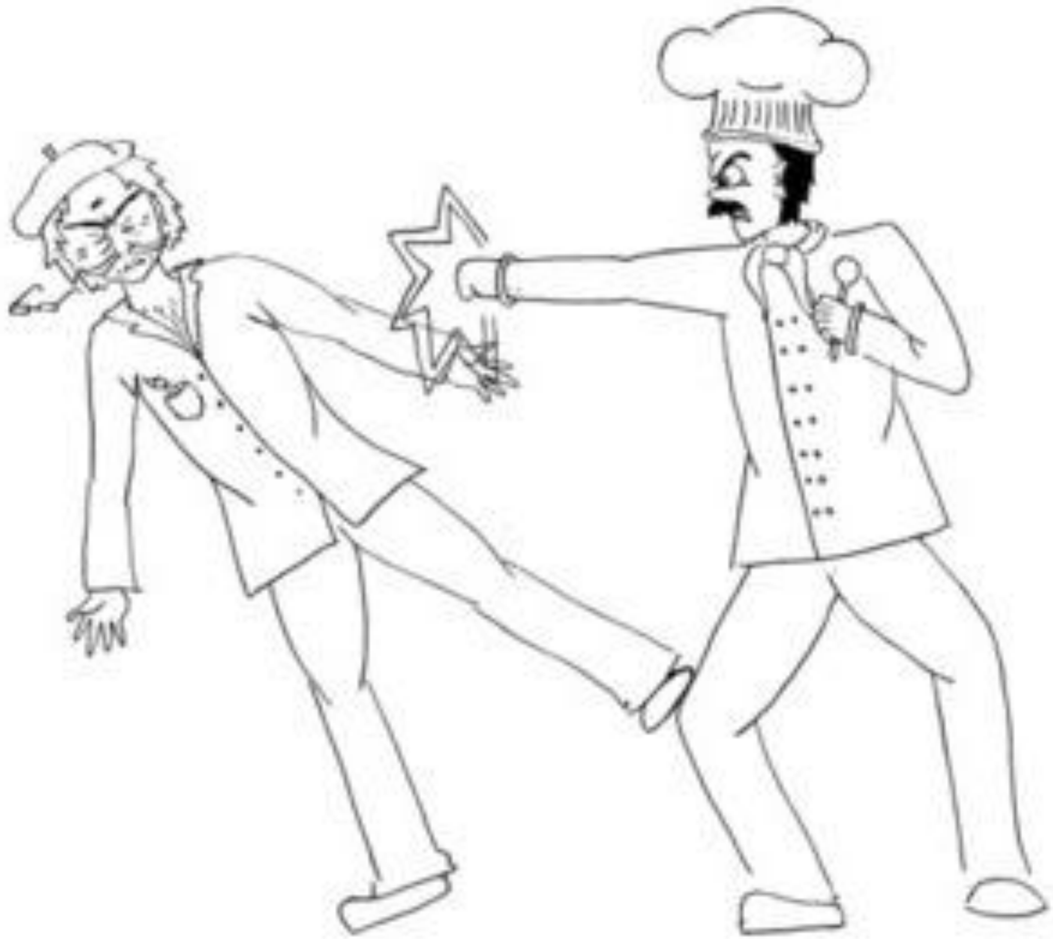


Information structure manipulation

→ painter ←

The painter punched the chef

} typical



Information structure manipulation

→ painter←

The painter punched the chef } typical

→ painter←

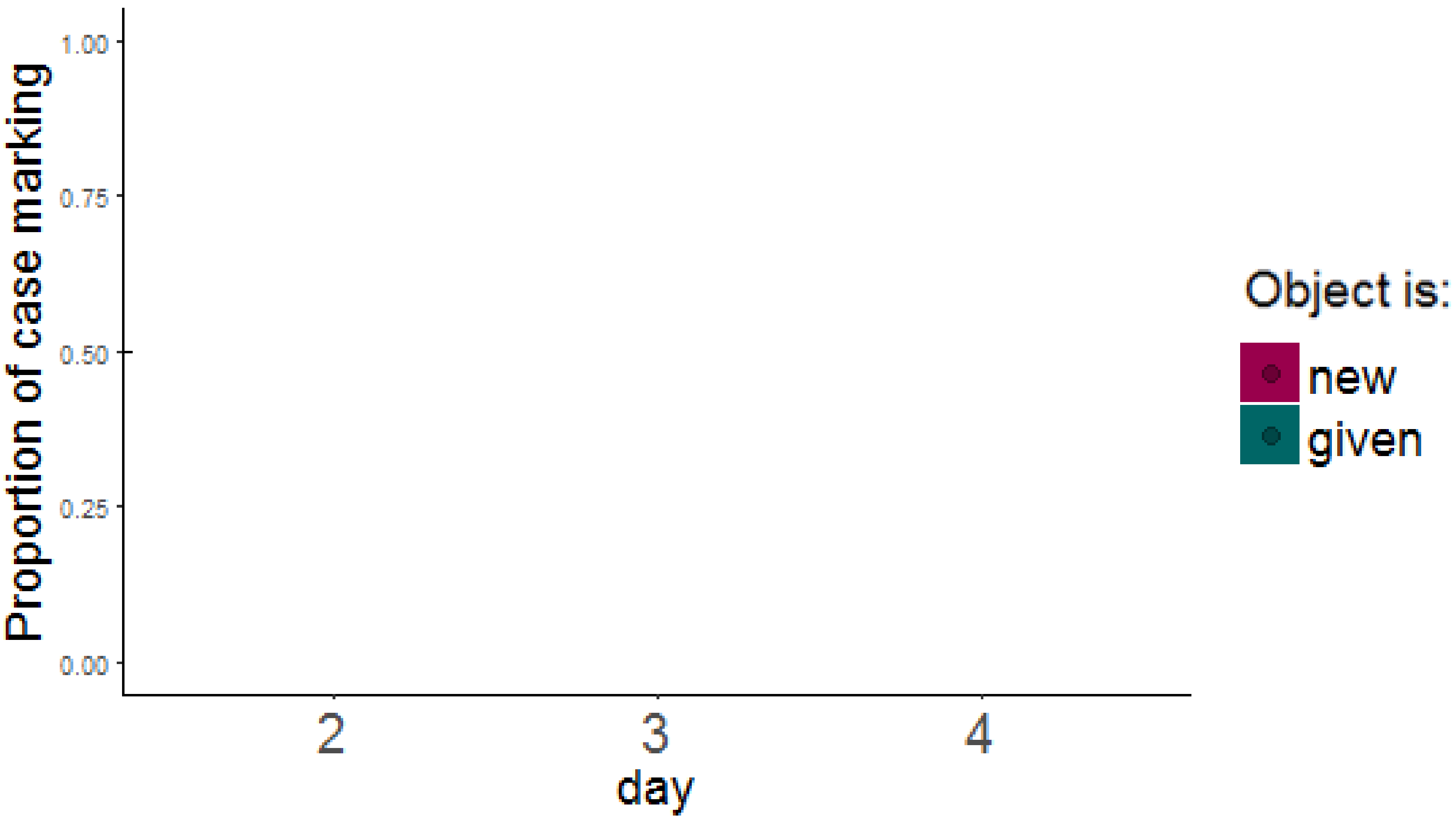
The chef punched the painter } atypical

Prediction

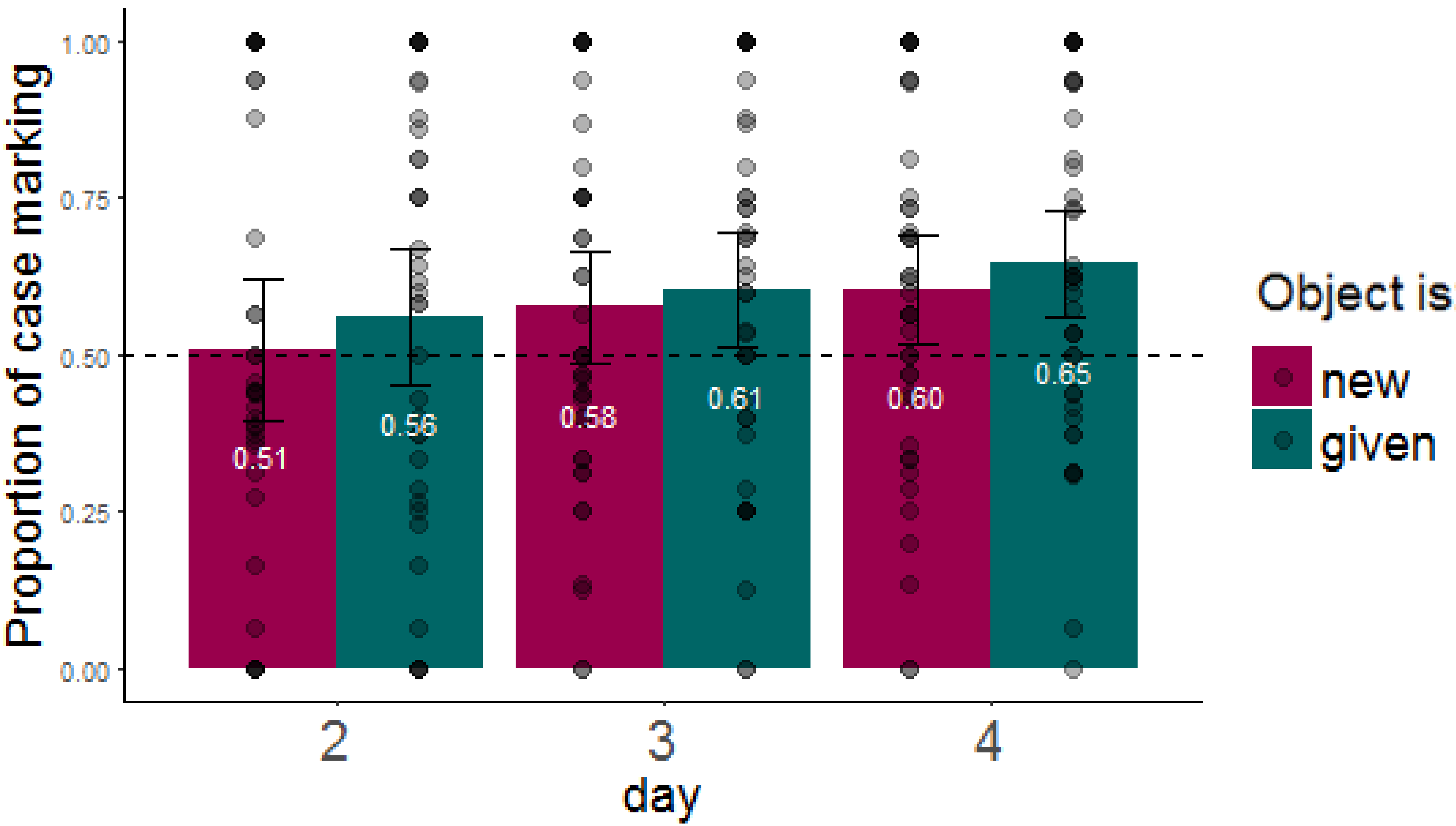
If case marking is impacted by information structure, speakers should shift their input to reflect that.

(increased marking for given objects)

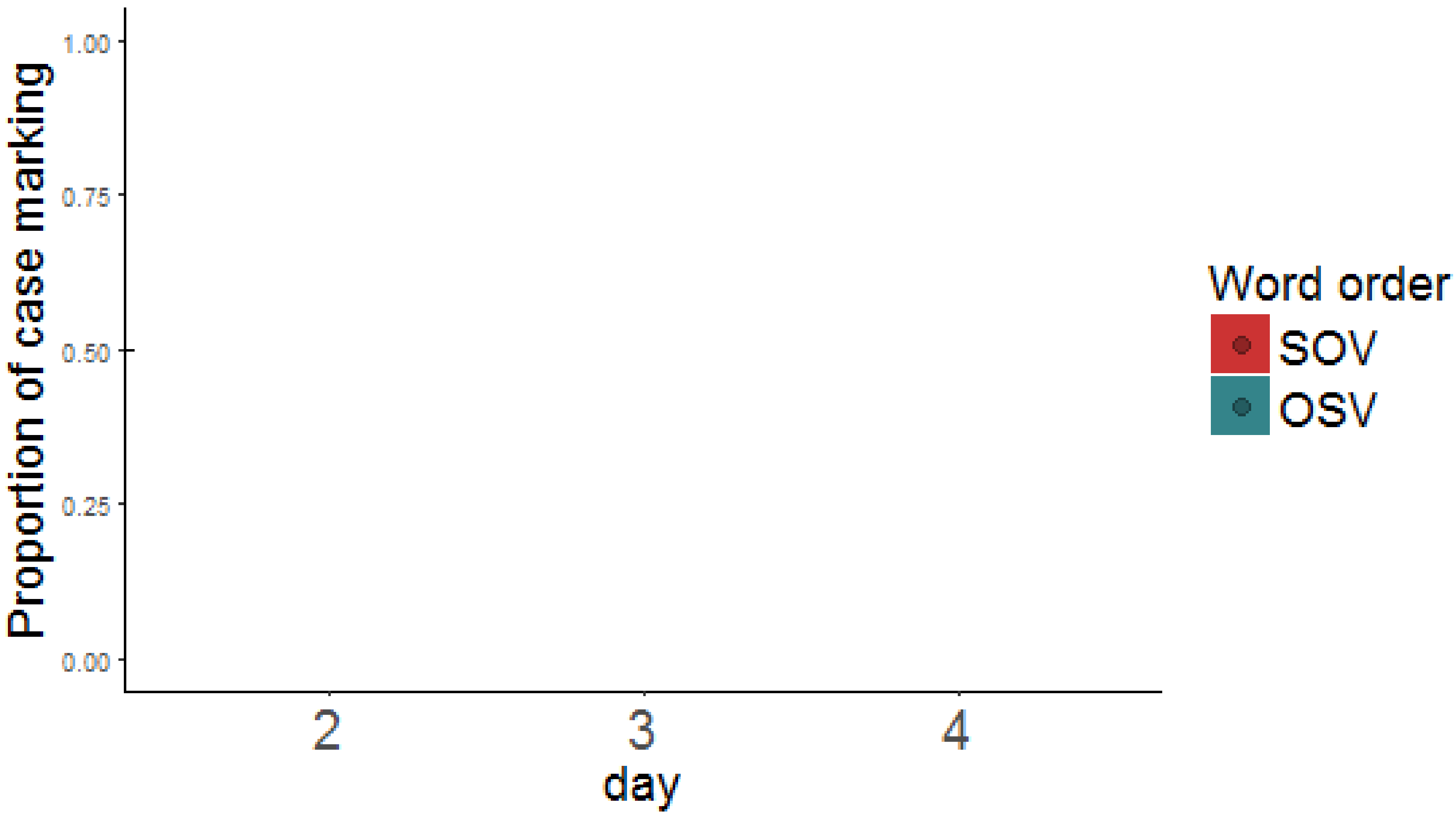
Did information structure impact case marking?



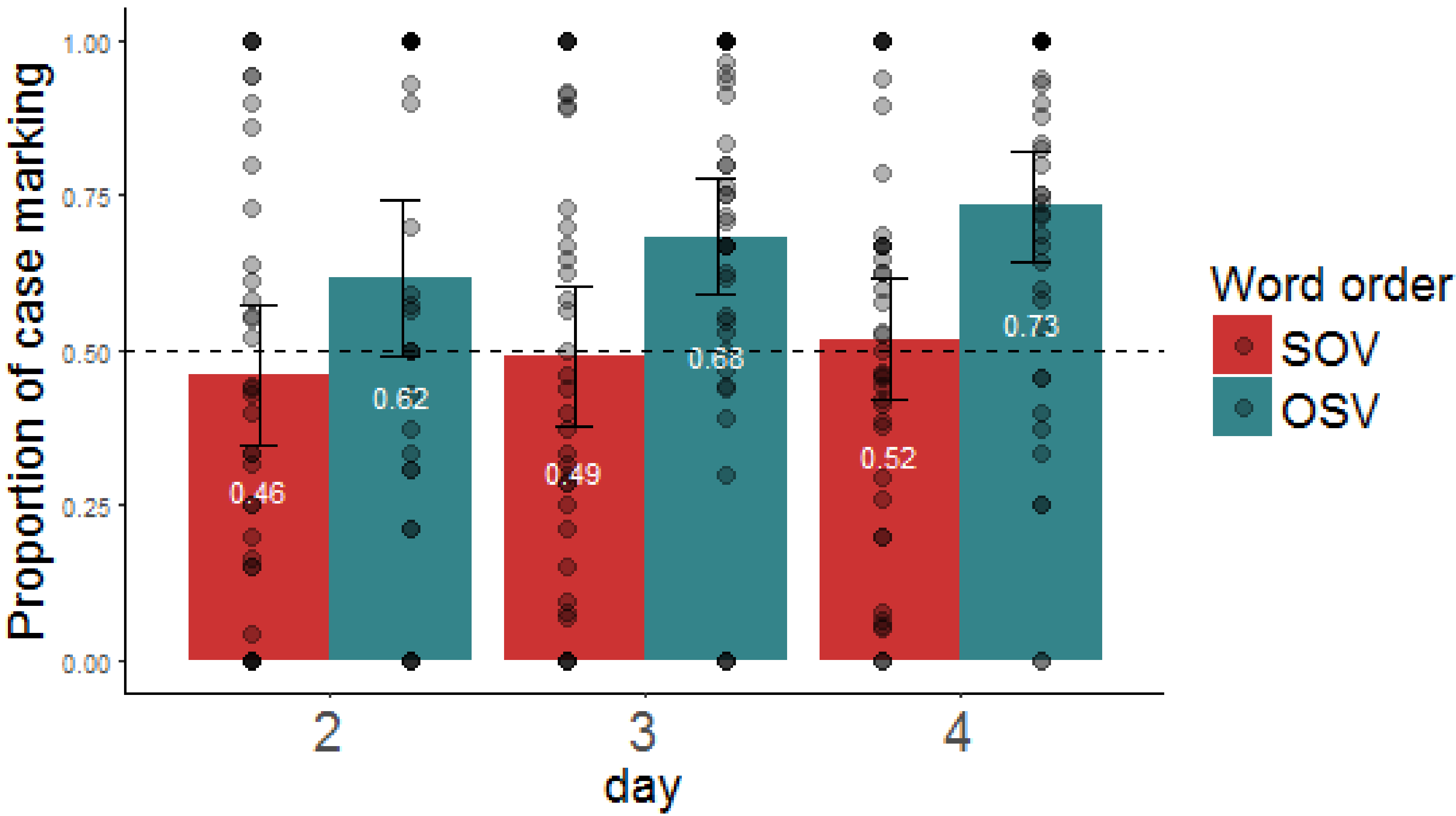
Proportion of case marking
1.00
0.75
0.50
0.25
0.00
2
3
4
day
Object is:
new
given



- No significant effect of information structure on case marking

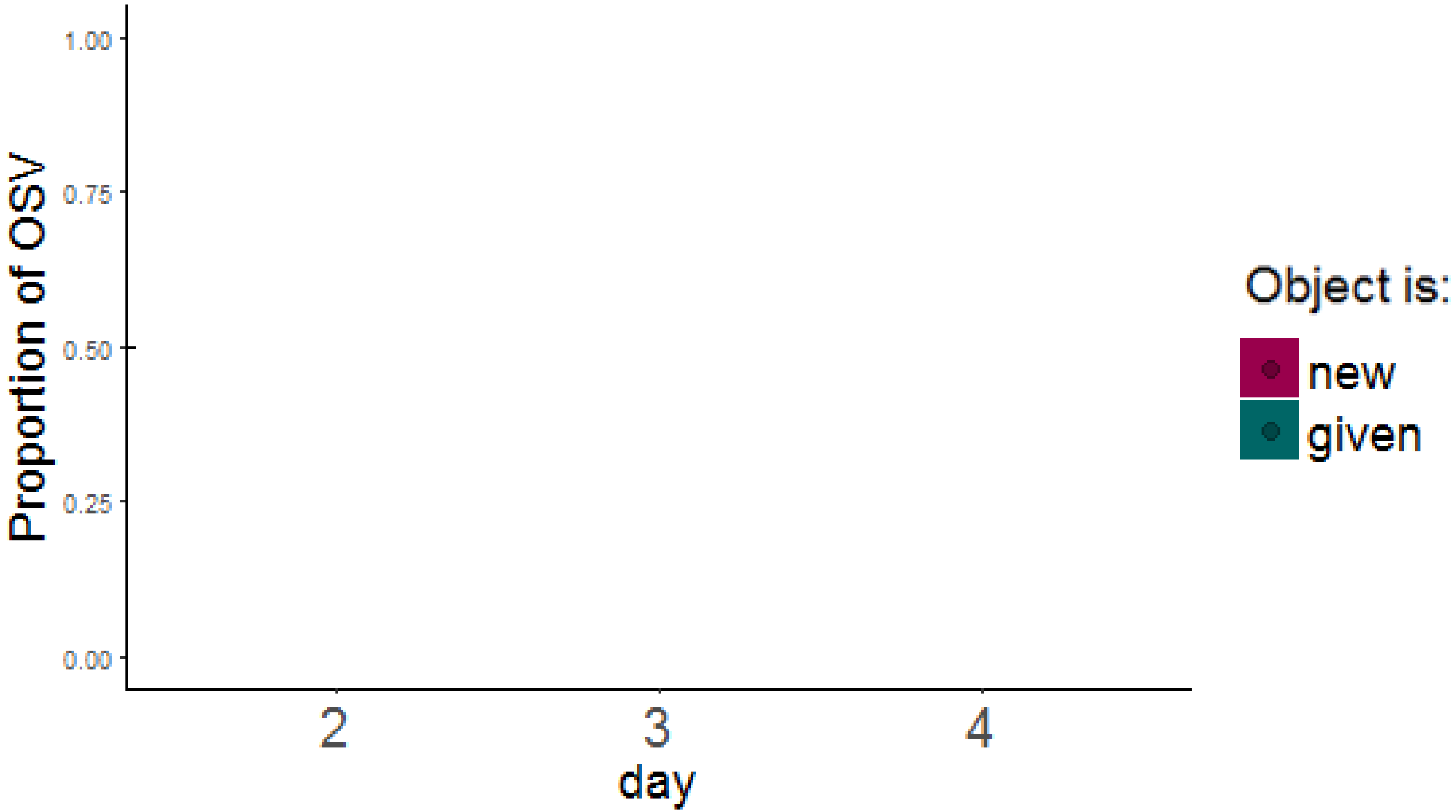


Proportion of case marking
1.00
0.75
0.50
0.25
0.00
2
3
4
day
Word order
SOV
OSV

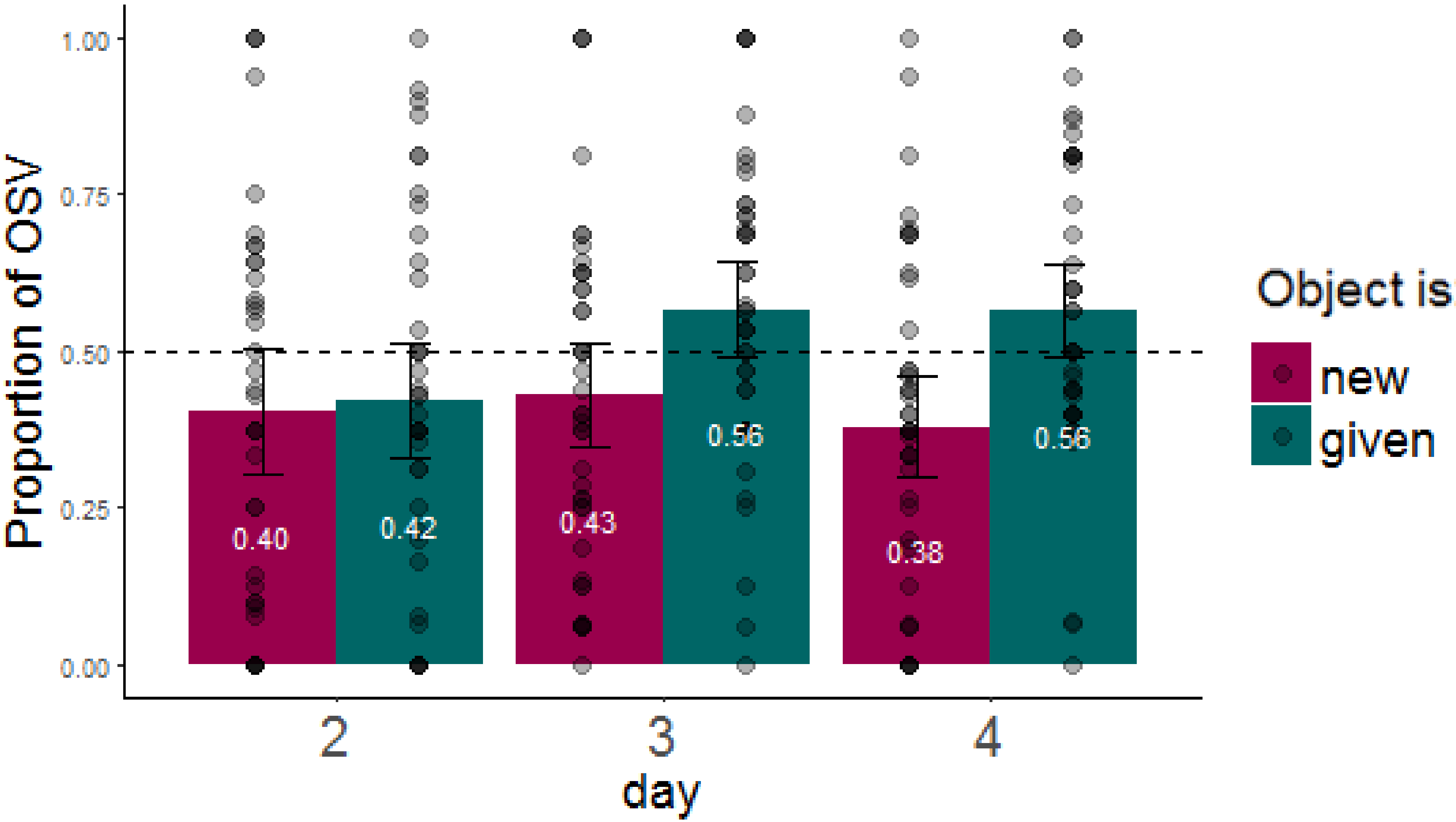


- Objects are case marked more in OSV

Did information structure impact word order?



1.00
0.75
0.50
0.25
0.00
Proportion of OSV
2
3
4
day
Object is:
new
given



• More OSV for given objects

Information structure and word order

- Word order is influenced by information structure (Tomlin, 1994; Arnold, Kaiser, Kahn & Kim, 2013)

- A general preference for given-before-new (for a review, see Arnold, Kaiser, Kahn & Kim, 2013; Kaiser & Trueswell, 2004)

What can we make of it?

- Indirect effect of givenness on case marking **through word order**

    - Word order was influenced by the manipulation

      given objects → OSV

    - Case marking was influenced by word order

      OSV → case marking

Evolution of DOM in natural languages

Evolution of DOM in natural languages

SUBJECT OBJECT VERB

- 1st stage:

  Given objects move to a “topical position”

(Downing, 2018 ; Iemmolo, 2010; 2011; 2013; Schurr, in prep)

Evolution of DOM in natural languages

OBJECT SUBJECT VERB

- 1st stage:

  Given objects move to a "topical position"

(Downing, 2018 ; Iemmolo, 2010; 2011; 2013; Schurr, in prep)

Evolution of DOM in natural languages

OBJECT+CASE SUBJECT VERB

- 1st stage:
  Given objects move to a “topical position”
- 2nd stage:
  These fronted constructions are case-marked
- 3rd stage:
  Expands to other properties of objects (e.g., animacy)

(Downing, 2018 ; Iemmolo, 2010; 2011; 2013; Schurr, in prep)

Evolution of DOM in natural languages

- The initial stages of the process can be simulated in the lab:
  - Given before new
  - Marking of atypical word order

- Next step: grammaticalization into semantic-based DOM?
  - E.g., creation of animacy-based case marking

Ongoing work

- Grammaticalization into semantic-based DOM
    - Can a manipulation of information structure bring about animacy-based marking?
- A (complicated) artificial language learning experiment
- No evidence so far (ask me more if you're interested!)

Cognitive biases and DOM

- Ambiguity avoidance (Jäger 2007; Fedzechkina, Jaeger & Newport, 2012, see also Gibson et al. 2013; Futrell et al. 2015)
  - Drives emergence of DOM **in the lab** (though see Smith & Culbertson, 2018)
  - May play a role in DOM **in natural language**
- Information structure (Rohlfs, 1971; Iemmolo, 2010; 2013)
  - Drives emergence of DOM **in natural languages**
  - May drive DOM via word order **in the lab**

- Different stages of one process?

THANKS!

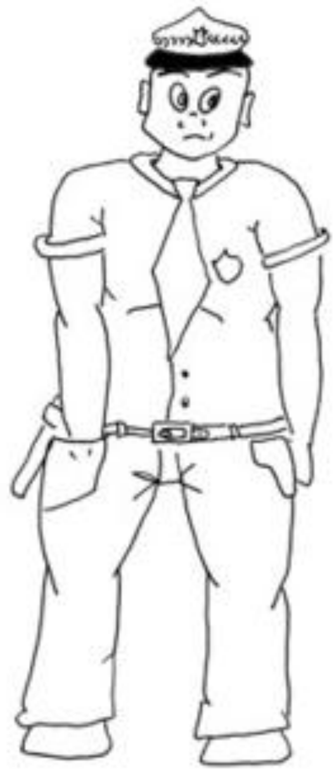
ERC grant 681942 to Kenny Smith

CENTRE for LANGUAGE EVOLUTION

Jack, Joseph and Morton Mandel School
for Advanced Studies in the Humanities

Procedure

1. NounTrain - 16 trials (each character twice).
2. NounComprehension - 8 trials.
3. SentenceTrain - 64 trials.
4. NounTrain - 16 trials (each character twice).
5. NounComprehension - 8 trials.
6. SentenceComprehension - 64 trials.
7. SentenceTest - 32 trials.



... click on the label that matches what you hear ... *(1/16)*

dacin

norg

Procedure

1. NounTrain - 16 trials (each character twice).
2. NounComprehension - 8 trials.
3. SentenceTrain - 64 trials.
4. NounTrain - 16 trials (each character twice).
5. NounComprehension - 8 trials.
6. SentenceComprehension - 64 trials.
7. SentenceTest - 32 trials.

... click on the choice that matches what you hear ...(1/8)

dacin

Procedure

1. NounTrain - 16 trials (each character twice).
2. NounComprehension - 8 trials.
3. SentenceTrain - 64 trials.
4. NounTrain - 16 trials (each character twice).
5. NounComprehension - 8 trials.
6. SentenceComprehension - 64 trials.
7. SentenceTest - 32 trials.

... click on the ***ACTION****...* *(3/64)*

lombur nagid prog

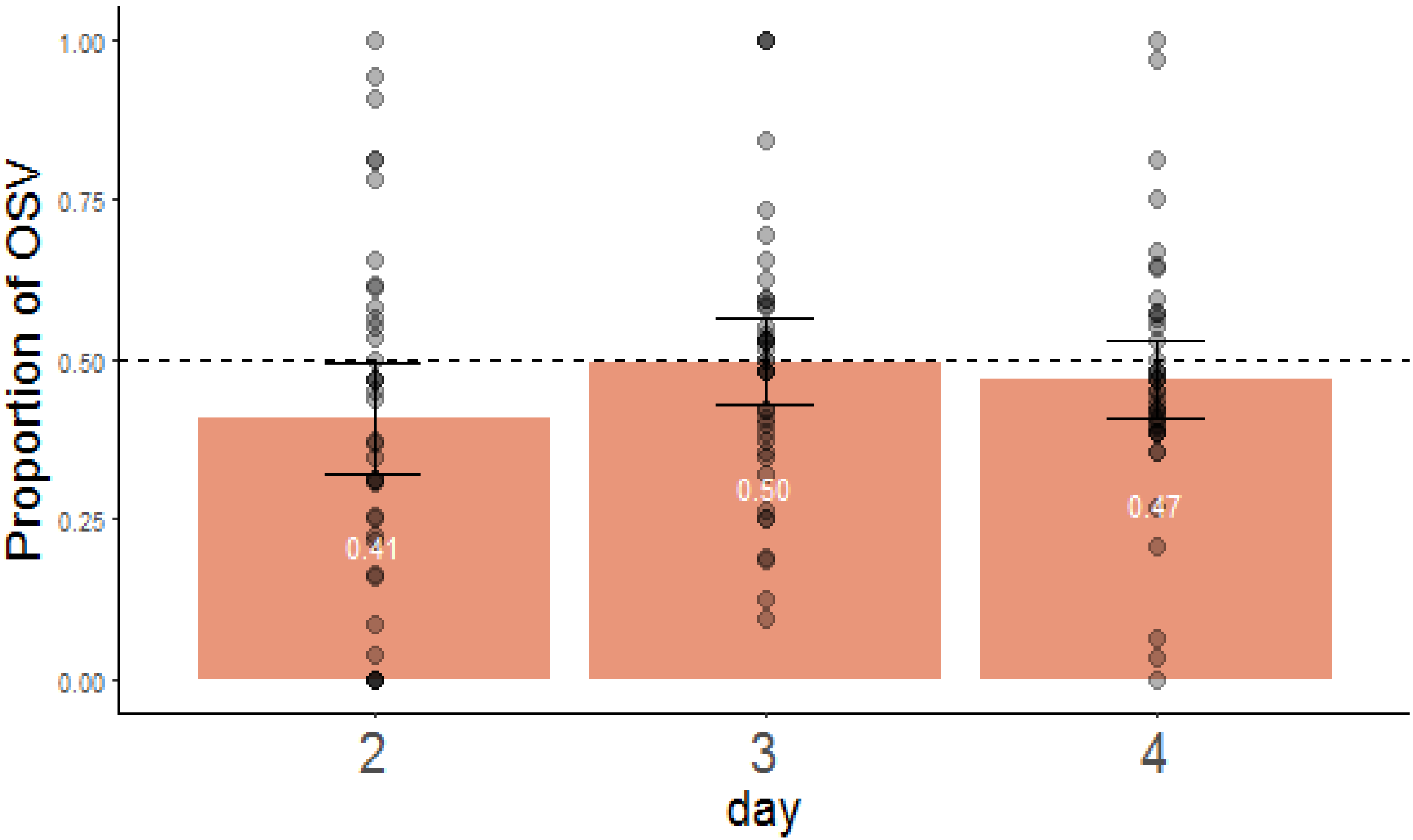
lombur	nagid	prog

Procedure

1. NounTrain - 16 trials (each character twice).
2. NounComprehension - 8 trials.
3. SentenceTrain - 64 trials.
4. NounTrain - 16 trials (each character twice).
5. NounComprehension - 8 trials.
6. SentenceComprehension - 64 trials.
7. SentenceTest - 32 trials.

Procedure

1. NounTrain - 16 trials (each character twice).
2. NounComprehension - 8 trials.
3. SentenceTrain - 64 trials.
4. NounTrain - 16 trials (each character twice).
5. NounComprehension - 8 trials.
6. SentenceComprehension - 64 trials.
7. SentenceTest - 32 trials.



- Participants start out disprefering OSV

Why mark OSV?

- General bias to put the agent first? (e.g., Bornkessel-Schlesewsky et al., 2008; Bickel et al., 2015)

- Effects from L1?
    - English speaking participants: expectations of S-initial sentences

- Deviations from expected associations tend to be coded by longer grammatical forms (Haspelmath, 2008; forthcoming)

Grammaticalization to animacy-based marking

Human > Animal > Inanimate

Definite > Indefinite

Given > New

Unmarked A
Marked P

Marked A
Unmarked P

(Bossong, 1991; Croft, 2003; Givón, 1976; Iemmolo, 2010; Silverstein, 1976)

Grammaticalization in the lab

Information structure	+	Animacy distinction	+	Correlating the two
Given objects → OSV OSV → case marking		Objects can be animates or inanimates (50%-50%)		Making animates more likely to be given than inanimates

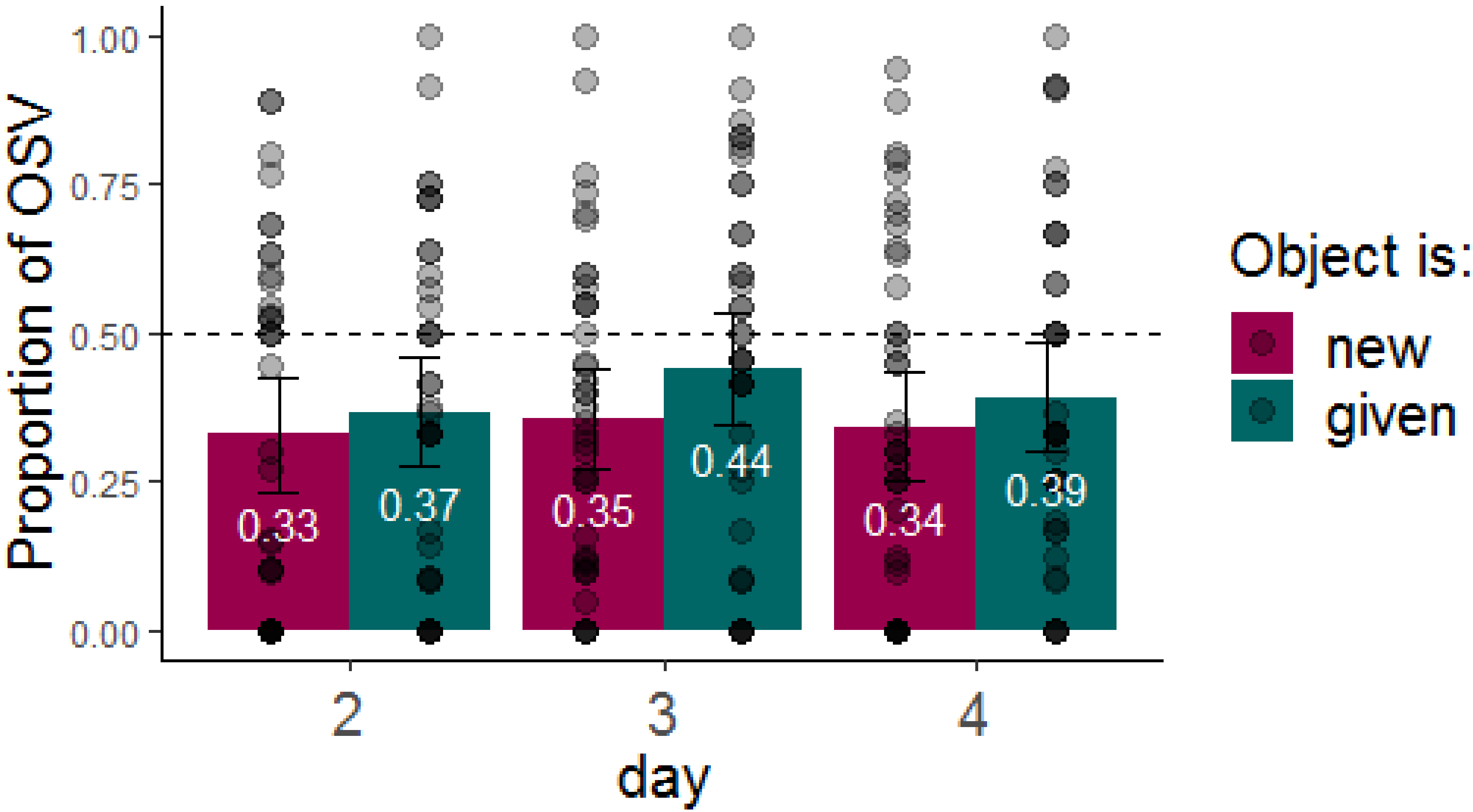
Together, this should bring about increased marking for animates

Grammaticalization experiment

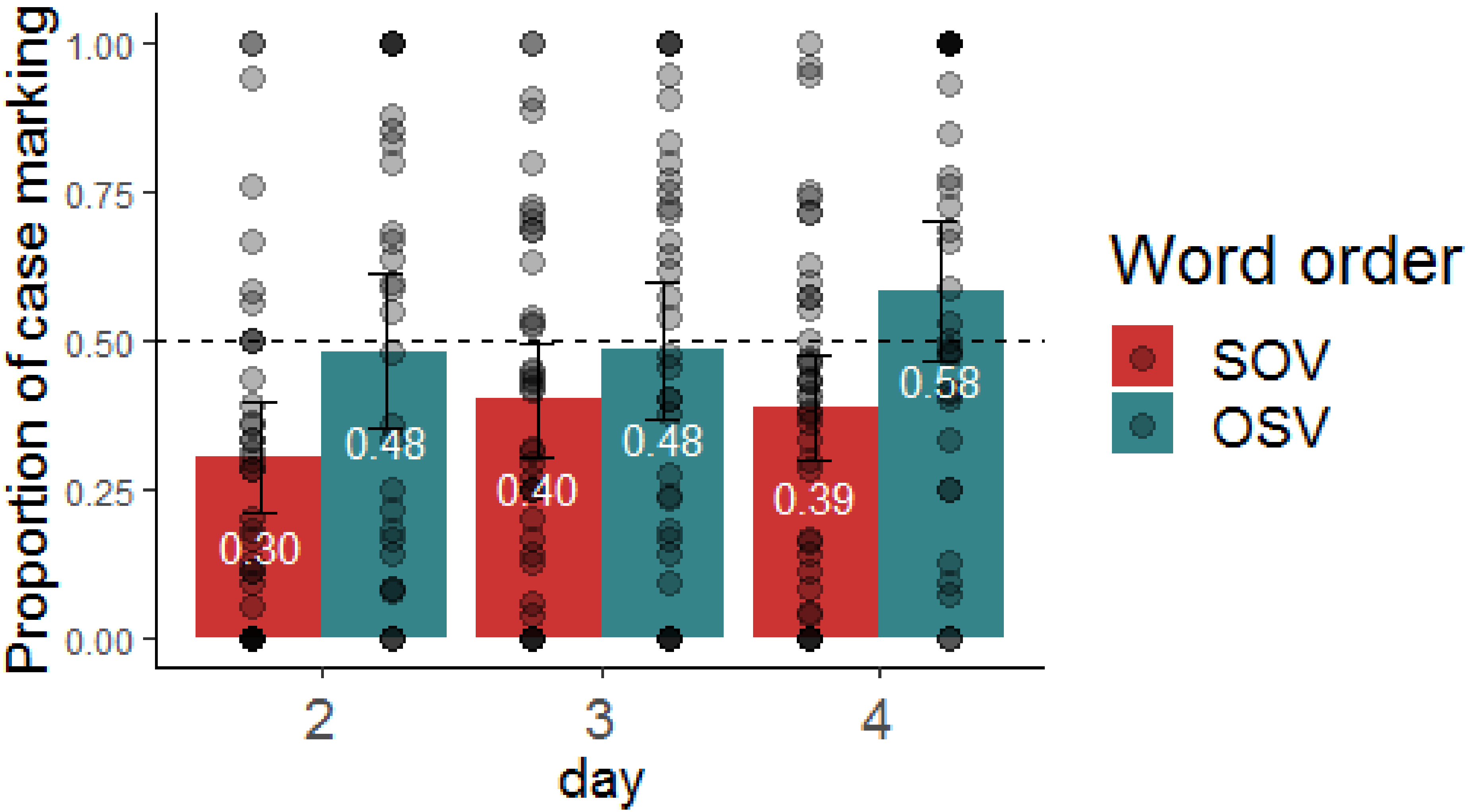
Same as Experiment1, but:

- Animacy distinction: Language has both animate and inanimate objects (50% each)
- Animates are more likely to be given than inanimates

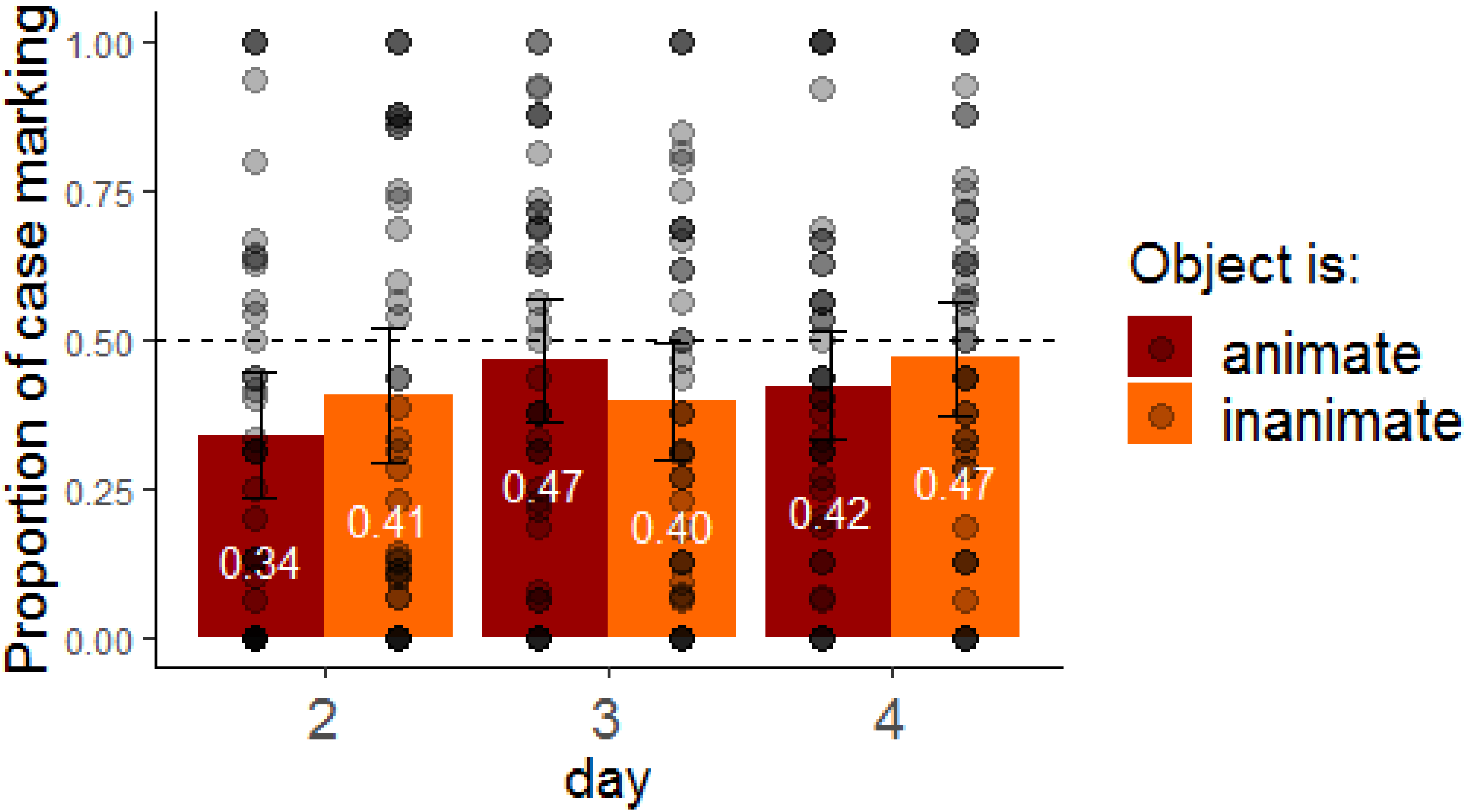
- Predictions:
  - given before new
  - increased marking for OSV
  - increased marking for animate objects (due to correlation between givenness and animacy)



Given before new



Increased marking for OSV



But no animacy effect…