**Definiteness across Languages: from German to Mandarin**

**David Bremmers, Jianan Liu, Martijn van der Klis & Bert Le Bruyn**

UIots, Utrecht University

We showcase the potential of a data-driven methodology for cross-linguistic research: Translation Mining (TM). We introduce the technique and put it to use in the domain of definiteness. We show how TM confirms existing insights about definiteness in German (Schwarz 2009) and Mandarin (Jenks 2018) while at the same time leading to novel insights for Mandarin.

## Introduction

- We showcase Translation Mining (TM). TM is a data-driven method that aims to confirm the relevance of established semantic dimensions while allowing to identify new ones. We focus on definiteness.
- Two dimensions of definiteness are well established in the literature: (i) uniqueness, and (ii) familiarity. English uses the for both.
- Schwarz (2009) argues German formally distinguishes between uniqueness/familiarity in the prepositional domain. Uniqueness definites contract with the preceding preposition (e.g. zum) whereas anaphoric definites do not (e.g. zu dem). Schwarz adopts the term WEAK for uniqueness definites and the term STRONG for anaphoric definites.
- Jenks (2018) is one of the cross-linguistic extensions of Schwarz’s seminal work. Jenks focuses on Mandarin and argues that it uses bare nouns for WEAK definiteness (see (1)) and demonstratives for STRONG definiteness (see (2)).

### (1) \( \text{Wäh/ Zhe } \text{ge} \) käum \( \text{zui} \) tief \( \text{hong} \) shangqì (bare noun)
- that/this CLF Taiwan’s president very angry
- 'The president of Taiwan is very angry'

### (2) Jiaoshi \( \text{li} \) zuó-zhe \( \text{yi} \) gē \( \text{nà} \) nánshèng \( \text{hè} \) \( \text{yì} \) \( \text{gē} \) nánshèng.
- classroom in sit-DUR one CLF boy and one CLF girl
- ‘Yesterday I met PRF that/this CLF boy
- ‘There are a boy and a girl sitting in the classroom. I met the boy yesterday.

### Translation Mining

- We use a parallel corpus consisting of Harry Potter and the Philosopher’s Stone and its translations to German and Mandarin. This corpus allows us to align markers of definiteness across languages in the same contexts.
- Our focus is on German and Mandarin because both have been argued to formally mark the WEAK/STRONG distinction. English is the perfect hub language: it uniformly relies on the and the definiteness dimensions we identify in German and Mandarin are consequently not influenced by translation.
- We extract German contracted and uncontracted PPs and align them with the English original and Mandarin translation.
- TM relies on semantic maps (henceforth TM maps) generated through multi-dimensional scaling (MDS) (Walchli & Cysouw, 2012; Behkhuizen et al., 2017; Van der Klis et al., 2017). The dots on these maps (e.g. Figure 1) stand for contexts from the corpus. MDS clusters contexts that use the same form in a given language and does so in one go for the three languages. This leads to a single clustering for the three languages. The colors of the dots are language specific and indicate which forms are used.
- The interpretation of TM maps is done both at the level of clusters and individual datapoints.

### Results

- The German TM map (Figure 1) displays neatly delineated clusters of blue and red dots. Blue stands for contexts with contracted PPs (N=40), red for contexts with uncontracted forms (N=56).
- Closer inspection of the contexts shows that the formal contracted/uncontracted distinction reflects the semantic WEAK/STRONG distinction (Bremmers 2019). Our results thus confirm Schwarz’ claim that German distinguishes between WEAK/STRONG definiteness in the prepositional domain:

#### (3) 'Ich denke, wir könnten ihn in den Zoo mitnehmen', sagte Tante Petunia langsам, und ihn im Wagen lassen... (contraped PP)
- I think we could him to the zoo take said aunt Petunia slowly... and him in the car leave
- ‘I supposed we could take him to the zoo,’ said Aunt Petunia slowly,... ‘and leave him in the car...’

#### (4) [Preceding discourse: “A giant of a man was standing in the doorway.”]
Harry sah zu dem Riesen auf. (uncontracted PP)
- Harry looked to the giant up
- ‘Harry looked up to the giant.’

### Discussion

- Jenks claims only confirm definiteness can be used for STRONG definiteness (see (2)). Our data go against this claim:

#### (5) [Preceding discourse: “As the owls flooded into the Great Hall as usual, everyone’s attention was caught at once by a long thin package carried by six large screech owls.”]
Tāmén pīshān-čè chībāng gānggāng fēi zōu. (Mandarin demonstrative)
- They flutter-DUR wings right fly away,
- ‘They had hardly fluttered out of the way when another owl dropped a letter on top of the parcel.’
- Our native speaker consultants confirm that (2) requires the demonstrative whereas (5) does not. We are thus facing two types of STRONG definite contexts.
- (5) is a regular narrative involving a sequence of events. (2) is different: the event of the sentence containing the anaphor does not follow the event of the sentence containing the antecedent. We hypothesize that this difference underlies the contrast in acceptability of the bare noun: bare nouns require their antecedents to be introduced in the same narrative sequence.
- Our intuition can be tested. We predict that turning (2) into a regular narrative sequence suffices to make the bare noun acceptable as an anaphoric expression.

### Conclusions

- TM confirms intuitions about the WEAK/STRONG distinction for German. It further confirms that Mandarin demonstratives can mark STRONG definiteness.
- TM identifies a new dimension within STRONG definiteness in Mandarin. This dimension is related to the distinction between narrative and non-narrative sequences.
- In Bremmers et al. (ms.), we explore how the narrative/non-narrative sequence distinction can be formalized through a dynamic interpretation of the situation variable in the analysis of STRONG definiteness in Schwarz (2009).
Semantic maps of German and Mandarin data

Figure 1. Semantic map showing German contracted and non-contracted definites.

Figure 2. Semantic map showing Mandarin equivalents of German data.

Sample of the corpus data

(1) German uncontracted → Mandarin demonstrative

E: But as all they knew for sure about the mysterious object was that it was about two inches long, they didn’t have much chance of guessing what it was without further clues.

G: Doch weil sie über das geheimnisvolle Ding nicht mehr wussten, als dass es though as they about the mysterious thing not more knew than that it gut fünf Zentimeter lang war, hatten sie ohne nähere Anhaltspunkte keine well five centimeters long was had they without further clues no große Chance zu erraten, was in dem Päckchen war.

M: Dànshì, guānyú nà gè shénmì wūjiàn, tāmén wéiyī nénggòu quèdìng de But about that CLF mysterious object they only can be sure NMLZ zhīshí tā-de chángdù yǒu liǎng yǐngcùn. Rúguò méi yǒu just its length have two inches if not have gèng duō de xiànsuǒ, shì bù kěnèng cǎi dào tā shì shénme more clue be not possible guess out it be what dōngxī de. thing NMLZ

Please visit https://time-in-translation.hum.uu.nl/ for more about Translation Mining and the Time in Translation project.
(2) **German uncontracted → Mandarin bare noun**

a. [**Preceding discourse:** As the owls flooded into the Great Hall as usual, everyone’s attention was caught at once by a long thin package carried by six large screech owls. Harry was just as interested as everyone else to see what was in this large parcel and was amazed when the owls soared down and dropped it right in front of him, knocking his bacon to the floor.]

**E:** They had hardly fluttered out of the way when another owl dropped a letter **on top of the parcel.**

**G:** Sie waren kaum aus dem Weg geflattert, als eine andere Eule einen Brief **auf**

they were hardly out the way fluttered, when a other owl a letter on

das Paket **warf.**

the package threw

**M:** Tāmen pūshan-zhe chìbǎng gānggāng fēi zǒu, yǒu yòu yì zhī māotóuyīng xié

they flutter-DUR wings right fly away, and have one CLF owl bring

lái yí fēng xīn, rěng zài bāoguó shàngmiàn.

come one CLF letter throw to parcel on.

deber spacious, zu der Schlange der Wanderer. … Das war Wundervolleres gekostet, doch immer noch konnte er den Blick nicht von dem Riesen abwenden.

**G:** Er reichte die Würstchen Harry, der so hungrig war, dass es ihm vorkam, als hätte er noch nie

b. [**Preceding discourse:** A giant of a man was standing in the doorway. …]

**E:** He passed the sausages to Harry, who was so hungry he had never tasted anything so wonderful, but he still couldn’t take his eyes off the giant.

**G:** Alle schwiegen und starrten hinauf **zu dem großen Schloss.**

**M:** Dājiā dōu chénmò wú yǔ, nǐngshí-zhe gāo rú yún tiān de jūdà chéngbāo.

everyone all silent no word stare-DUR high to cloud sky NMLZ huge castle

c. [**Preceding discourse:** The narrow path had opened suddenly on to the edge of a great black lake. Perched atop a high mountain on the other side, its windows sparkling in the starry sky, was a vast castle with many turrets and towers.]

**E:** Everyone was silent, staring up **at the great castle** overhead.

**G:** Du besuchten und näherten sich **zu dem Schloss.**

**M:** Tā huí guò tóu lǎi kān-zhe jù māng, yē duì tā zhā le zhā yǎn.

he turn back head come look-DUR big snake also to it wink PFV wink eye

d. [**Preceding discourse:** After lunch they went to the reptile house. It was cool and dark in here, with lit windows all along the walls. … Dudley quickly found the largest snake in the place….Harry moved in front of the tank and looked intently at the snake. … The snake suddenly opened its beady eyes. Slowly, very slowly, it raised its head until its eyes were on a level with Harry’s. It winked.]

**E:** He looked back at the snake and winked, too.

**G:** Er drehte sich wieder **zu der Schlange um und zwinkerte zurück.**

**M:** Tā huí guò tóu lǎi kān-zhe jù māng, yē duì tā zhā le zhā yǎn.

he turn back head come look-DUR big snake also to it wink PFV wink eye

e. [**Preceding discourse:** ‘Caput Draconis,’ said Percy, and the portrait swung forward to reveal a round hole in the wall. … ‘Come on,’ he said to Ron. He pushed open the portrait of the Fat Lady and climbed through the hole. Hermione wasn’t going to give up that easily.]
E: She followed Ron through the portrait hole, hissing at them like an angry goose.

G: Sie folgte Ron durch das Loch hinter dem Bild und fauchte wie eine wütende Gans.

M: 她跟著倫爬進洞口，像只發怒的鴨子一樣，對著他們嘶吼。

(3) German contracted → Mandarin demonstrative

E: ‘I’m not having one in the house, Petunia!’

G: ‘Ich will keinen davon im Haus haben, Petunia!’

M: ‘佩妮，我們不准讓任何人進這個房子。’

Petunia I not have them anyone enter this house

(4) German contracted → Mandarin bare noun

E: While Uncle Vernon made furious telephone calls to the post office and the dairy trying to find someone to complain to, Aunt Petunia shredded the letters in her food mixer.

G: Während Onkel Vernon wütend beim Postamt und bei der Molkerei anrief und versuchte jemanden aufzutreiben, bei dem er sich beschweren konnte, zerschnittete Tante Petunia die Briefe in ihrem Küchenmixer.

M: 芳妮用碎布把信封剪成碎片。

Petunia Aunt the letters in her blender

(5) Mandarin variation on (2)

M: 麦格教授前一日晚已经寄了一封信给哈利 #（这封）包裹。

‘Professor McGonagall had sent the package the day before.’

Notes:

1. The subject position in Mandarin is special in that bare nouns can be used as STRONG definites. We sidestep the complication of the subject/non-subject position.

2. We do not focus on German prepositions in general, but only look at those which can have contracted forms with the following articles.
References
Heim, I. 1982. The semantics of definite and indefinite noun phrases.