Definiteness across Languages: from German to Mandarin

David Bremmers, Jianan Liu, Martijn van der Klis & Bert Le Bruyn
UiL OTS, 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) (*#Nà /#Zhè gè) táiwān de **zǒngtǒng** hěn shēngqì* **(bare noun)**
that/this CLF Taiwan 's president very angry
'The **president** of Taiwan is very angry'

(2) *Jiàoshì lǐ zuò-zhe yī gè **nánshēng** hé yī gè nǚshēng,*
classroom in sit-DUR one CLF boy and one CLF girl
*Wǒ zuótiān yùdào le **#(nà/ zhè gè)** **nánshēng*** **(demonstrative)**
I yesterday meet 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) (Wälchli & Cysouw, 2012; Beekhuizen 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).

Figure 1. TM map of German definiteness.

- 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 langsam,*
I think we could him to the zoo take said aunt Petunia slowly
*... und ihn **im Wagen** lassen ...'* **(contracted PP)**
and him in the car leave
"'I suppose 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**."

- For German, **TM confirms existing intuitions** about the WEAK/STRONG dimension. This outcome allows us to establish the **German** contracted/uncontracted clusters **as a baseline** for exploring the WEAK/STRONG dimension in Mandarin.
- With German as a baseline, Jenks (2018) leads us to expect the clusters of **Mandarin** bare noun and demonstrative contexts to be identical to the clusters we find for contracted and uncontracted PPs in German. This is not what we find. The TM maps in Figures 2 and 3 illustrate. Frequencies are given below.

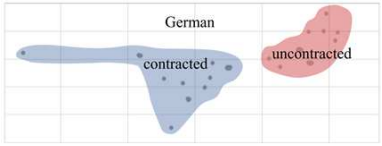

Figure 2. TM map of German definiteness.

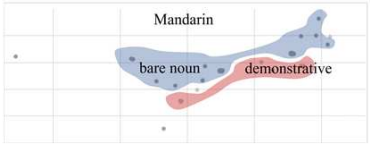

Figure 3. TM map for Mandarin definiteness.

German contracted form (N=40) (*e.g.* (3)) → Mandarin demonstrative (N=3)
→ Mandarin bare noun (N=34)
German uncontracted form (N=56) (*e.g.* (4)) → Mandarin demonstrative (N=10)
→ Mandarin bare noun (N=45) (*e.g.* (5))

- Closer inspection of the contexts reveals two things: (i) **Mandarin demonstratives** appearing as equivalents of German contracted forms (N=3) **do not function as WEAK definites** but get a deictic interpretation. (ii) **Mandarin bare nouns** appearing as equivalents of German uncontracted forms (N=45), however, do **get a STRONG definite interpretation**.

Discussion

- Jenks claims only demonstratives 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."]
Tāmen pūshan-zhe chìbǎng gānggāng fēi zǒu,
They flutter-DUR wings right fly away,
yòu yǒu yì zhī māotóuyīng xié lái yì fēng xìn,
and have one CLF owl bring come one CLF letter
*rēng zài **bāoguǒ** shàngmiàn.* **(bare noun)**
throw to parcel on.
'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. (6) shows that this prediction is borne out.

(6) *Jiàoshì lǐ zuò-zhe yí gè **nánshēng** hé yí gè nǚshēng,*
classroom in sit-DUR one CLF boy and one CLF girl
*wǒ jìn jiàoshì dǎ le **nánshēng.*** **(bare noun)**
I enter classroom hit PRF boy
'There were a boy and a girl sitting in the classroom. I entered and hit **the boy**.'

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).

Scan the QR code to visit the website https://time-in-translation.hum.uu.nl/ for more about Translation Mining and the Time in Translation project.

Definiteness across Languages: from German to Mandarin

David Bremmers, Jianan Liu, Martijn van der Klis & Bert Le Bruyn
UiL OTS, Utrecht University

Semantic maps of German and Mandarin data

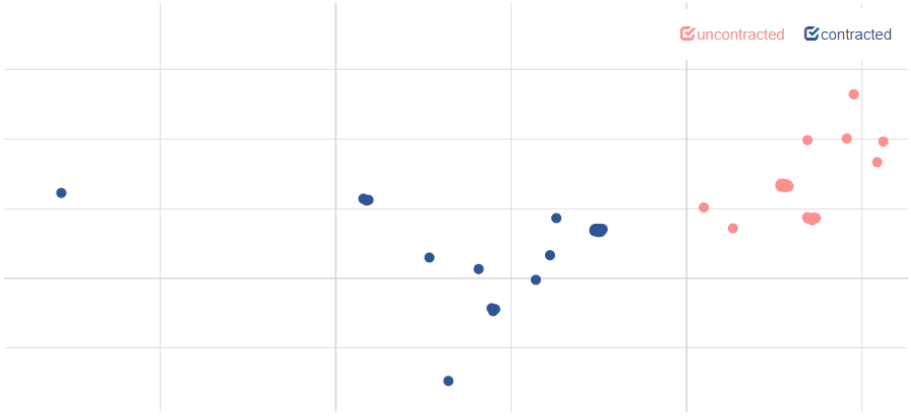

Figure 1. Semantic map showing German contracted and non-contracted definites.

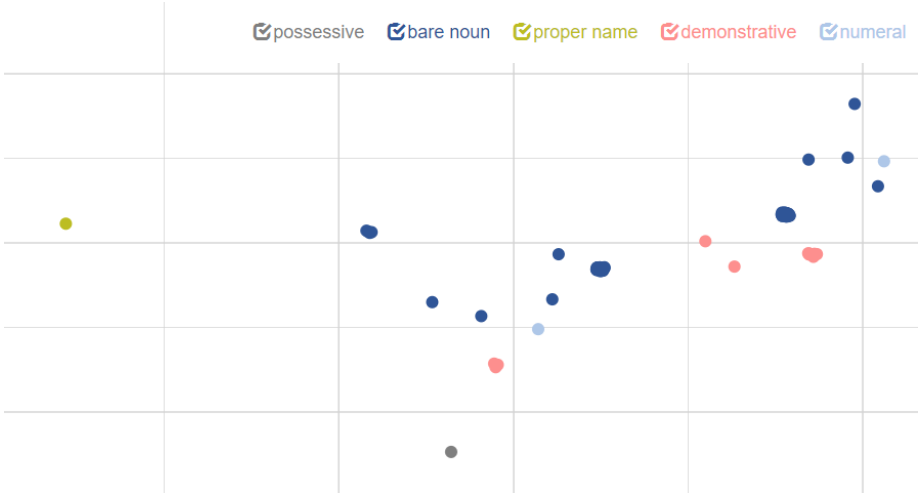

Figure 2. Semantic map showing Mandarin equivalents of German data.

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.
big chance to guess what in the package was

M: *Dànshì, guānyú* ***nà gè shénmì wùjiàn****, tāmen wéiyī nénggòu quèdìng de*
But about that CLF mysterious object they only can be sure NMLZ
zhǐshì tā-de chángdù yǒu liǎng yīngcùn. Rúguǒ méi yǒu
just its length have two inches if not have
gèng duō de xiànsuǒ, shì bù kěnéng cāi dào tā shì shénme
more clue be not possible guess out it be what
dōngxī de.
thing NMLZ

(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, yòu yǒu yì zhī māotóuyīng xié*
they flutter-DUR wings right fly away, and have one CLF owl bring
lái yì fēng xìn, rēng zài ***bāoguǒ*** *shàngmiàn.*
come one CLF letter throw to parcel on.

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: *Er reichte die Würstchen Harry, der so hungrig war, dass es ihm vorkam, als hätte er noch nie etwas Wundervolleres gekostet, doch immer noch konnte er den Blick nicht* ***von dem Riesen*** *abwenden.*

M: *Tā bǎ xiāngcháng dì gěi hā lì, hā lì zǎojiù è jí le.*
he BA sausage pass to Harry Harry already hungry extreme PFV
Tā zhè bèizi yěméi chī guò zhème hàochī de dōngxī, Dàn tā
he this life never eat EXP this delicious thing But he
shǐzhōng wúfǎ jiāng mùguāng cóng ***jùrén*** *shēn shang yí kāi.*
from the start to the end cannot AUX eyesight from giant body on move away

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

M: *Dàjiā dōu chénmò wú yǔ níngshì-zhe gāo rù yún tiān de* ***jùdà chéngbǎo***.
everyone all silent no word stare-DUR high to cloud sky NMLZ huge castle

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 lái kàn-zhe* ***jù mǎng***, *yě duì tā zhǎ le zhǎ 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: *Tā gēn-zhe luōēn pá jìn **dòngkǒu**, xiàng yì zhī fānù de mǔ é yādī shēngyīn*
she follow-DUR Ron climb in hole like one CLF angry goose lower sound
cháo tāmen rāngrang.
at them hiss

(3) German contracted → Mandarin demonstrative
E: 'I'm not having one **in the house**, Petunia!'
G: '*Ich will keinen davon **im Haus** haben, Petunia!*'
I want none of.those in.the house have Petunia
M: '*Pèinī, wǒ juébù ràng tāmen rènhérén jìn **zhè dòng fángzi**.*'
Petunia I not have them anyone enter this CLF 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, zerschnitzelte Tante Petunia die Briefe in ihrem Küchenmixer.*
while uncle Vernon angrily at.the post.office and at the dairy called and tried someone to.find at who he himself complain could shredded aunt Petunia the letters in her blender
M: *Fèi nóng yífu nù chōngchōng de gěi **yóujú**, nǎi chǎng*
Vernon Uncle furiously to post office dairy
dǎ diànhuà zhǎo rén shuōlǐ. Pèi nī yímā zhènghǎo bǎ
make call find someone complain Petunia Aunt just BA
èrshísì fēng xìn dōu sāi dào shípǐn fěnsuì jī lǐ jiǎo de fěnsuì
twenty-four CLF letter all insert to food processor in shred to pieces

(5) Mandarin variation on (2)
[**Preceding discourse**: same as (2)]
M: *Màigé jiàoshòu qián yī tiān jì gěi hālì #(**zhè ge**) **bāoguǒ**.*
McGonagall Professor before one day send to Harry this CLF package
'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

Beekhuizen, B., J. Watson & S. Stevenson. 2017. Semantic Typology and Parallel Corpora: Something about Indefinite Pronouns. In *CogSci*.

Bremmers, D. 2019. La définitude en français, anglais, allemand et mandarin: Les références définies dans les contextes prépositionnels. (Bachelor thesis).

Bremmers, D., J. Liu, M. van der Klis & B. Le Bruyn. 2019. Translation Mining: definiteness across languages. A reply to Jenks (2018). Manuscript, Utrecht University. Available at https://time-in-translation.hum.uu.nl/.

Jenks, P. 2018. Articulated definiteness without articles. *Linguistic Inquiry*. Volume 49, 3:501- 536.

Heim, I. 1982. *The semantics of definite and indefinite noun phrases.*

Kamp, H.1981. A theory of truth and semantic representation. In *Formal Methods in the Study of Language*, edited by Jeroen A. G. Groenendijk, T. M. V. Janssen, and Martin B. J. Stokhof, 277-322. Mathematical Center Tract 135, Amsterdam.

Russell, B. 1905. On denoting. *Mind* 14:479-493

Strawson, P. 1950. On referring. *Mind* 59,235: 320-344.

Schwarz, F. 2009. *Two Types of Definites in Natural Language*. Amherst, Massachusetts: University of Massachusetts Amherst dissertation.

Van der Klis, M., B. Le Bruyn & H. de Swart. 2017. Mapping the perfect via translation mining. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 497-502.

Wälchli, B. & M. Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3), 671-710.