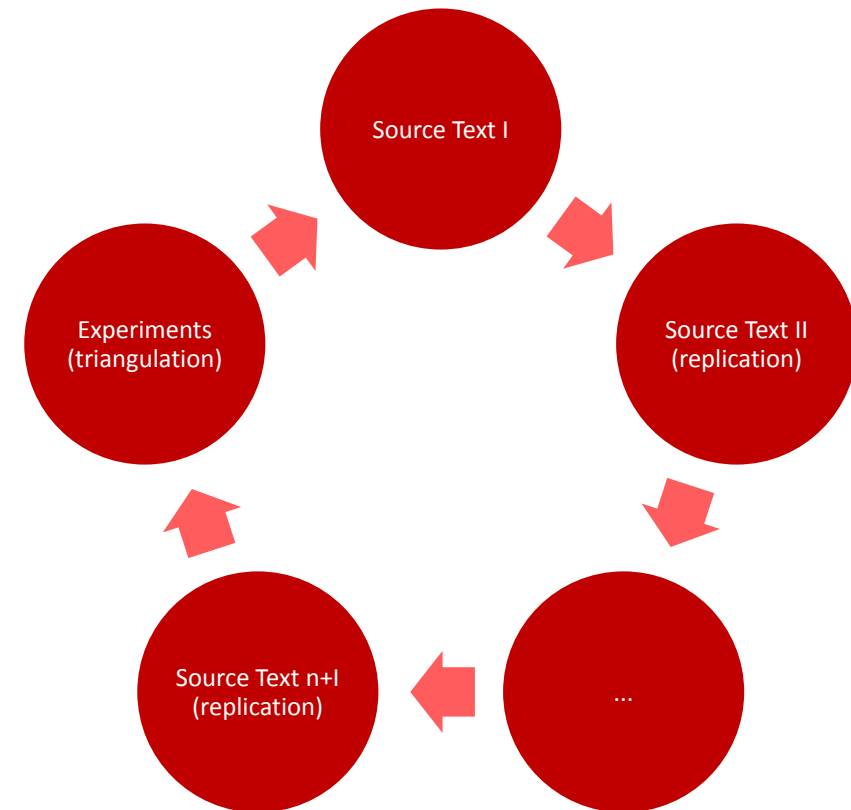

Reference: *a Translation Mining view*



Translation Mining

Translation Mining

- > ° Time in Translation project (Henriëtte de Swart & Bert Le Bruyn | Utrecht | 2017-2022).
- > Exploiting translations to map out and analyze cross-linguistic variation.
- > Research method complementing questionnaires, experiments, monolingual corpus research, etc.
- > Methodological reflection (see, e.g., Le Bruyn et al. 2023; Le Bruyn et al. 2024).



Translation Mining

- > Original focus on **meso-variation** in the use of **HAVE/BE + past participle** in Western European languages.
- > Main methodological insights:
 - Translation data are **highly systematic**, allowing for insightful analysis of general patterns of variation and of the individual datapoints that give rise to them;
 - They are particularly useful in the study of the **division of labor between forms**, within and across languages;
 - They can be used both from an **exploratory** and a **hypothesis-driven** perspective.

Translation Mining

Fragment overview



Source

French

1.xml

Passé Composé



↻ s5.14 J' ai lu le dossier de votre mère .

1

Translations

German

Perfekt



Ich habe die Akte Ihrer Mutter gelesen .

English

Present Perfect



I've read your mother 's file .

Spanish

Pretérito Perfecto Compuesto



He leído el expediente de su madre .

Italian

Passato Prossimo



Ho letto la pratica di vostra madre .

Dutch

Vtt



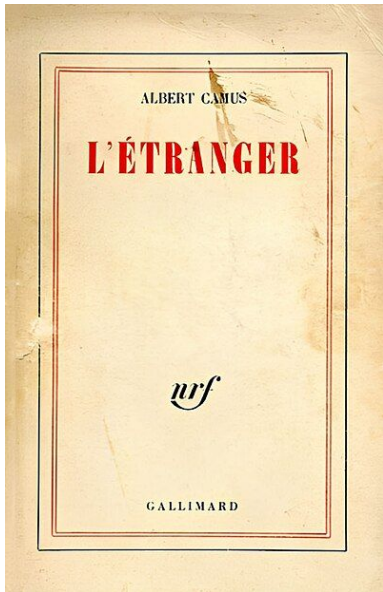
Ik heb uw moeders map doorgelezen .

Portuguese

Pretérito Perfeito Simples

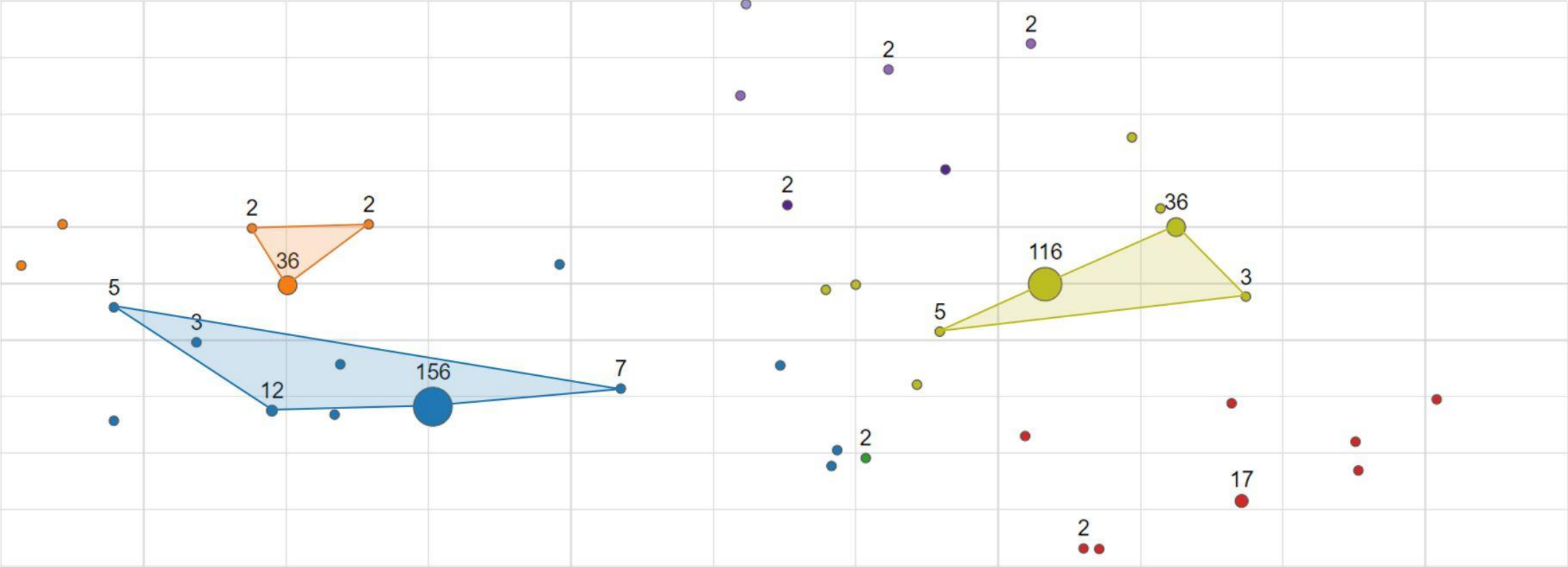


Estive lendo o processo da sua mãe .



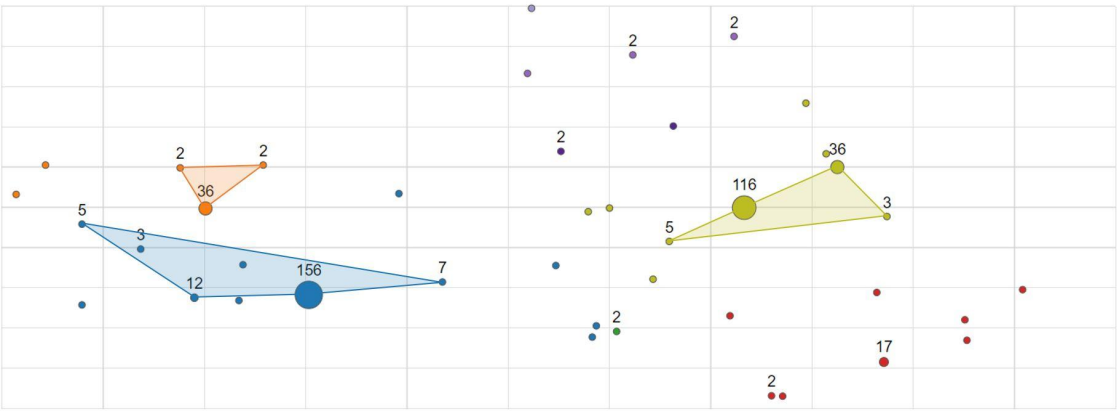
Translation Mining

passé composé présent futur simple imparfait futur antérieur plus-que-parfait futur proche passé simple

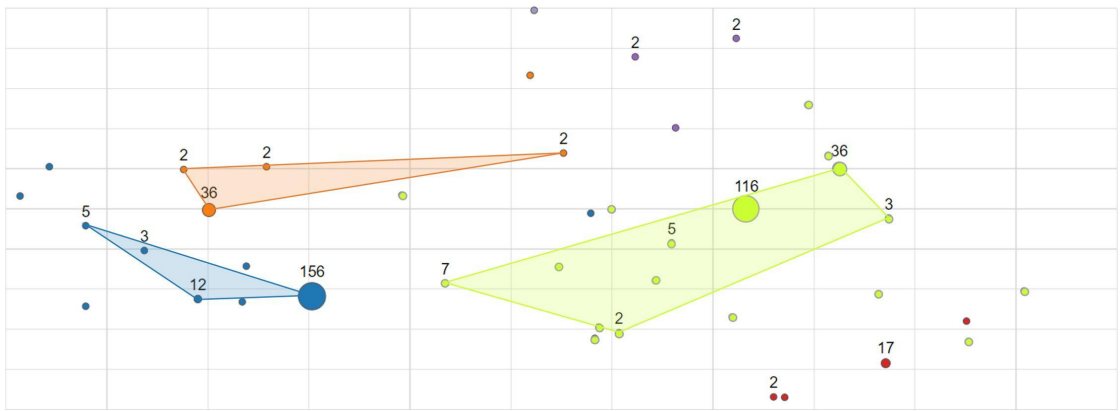


Translation Mining

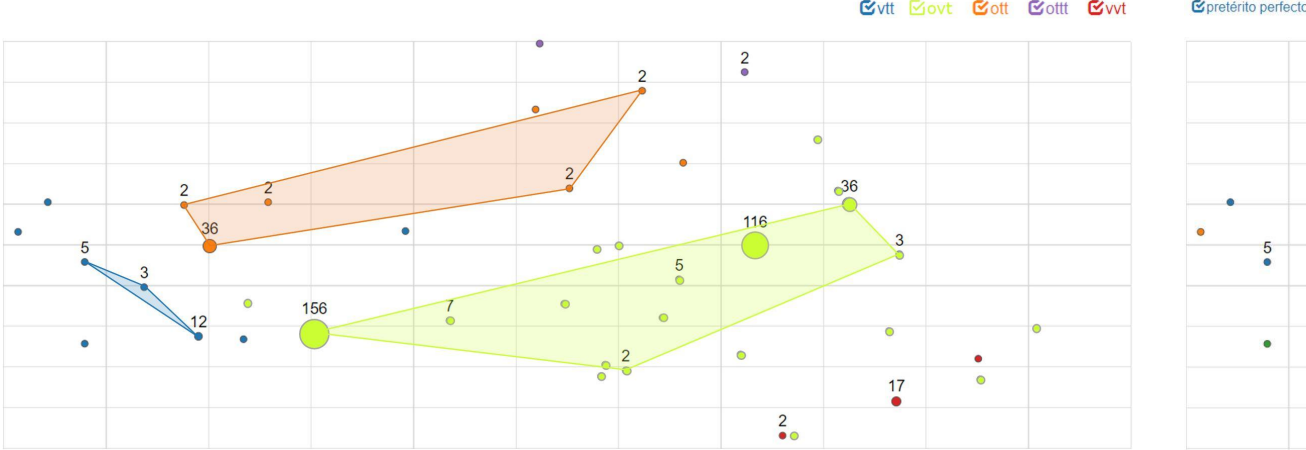
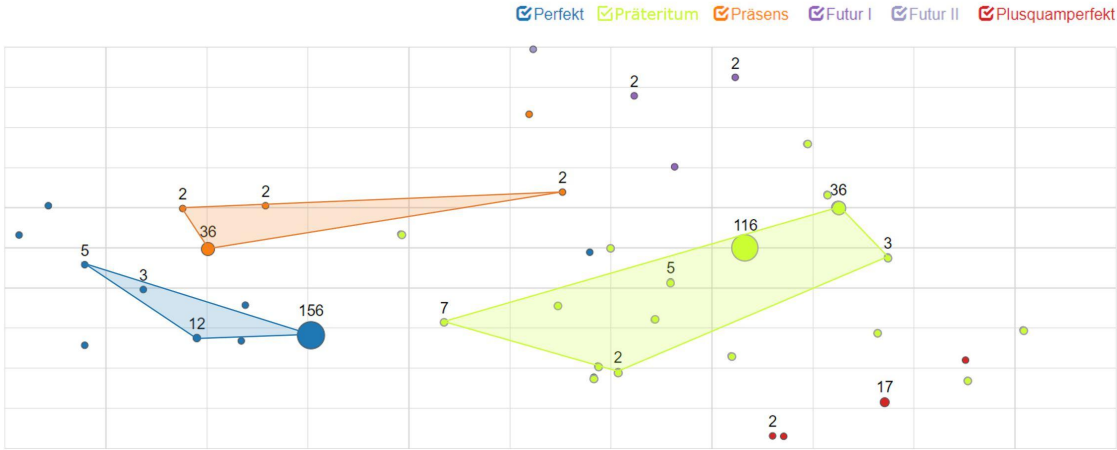
passé composé
 présent
 futur simple
 imparfait
 futur antérieur
 plus-que-parfait
 futur proche
 passé simple



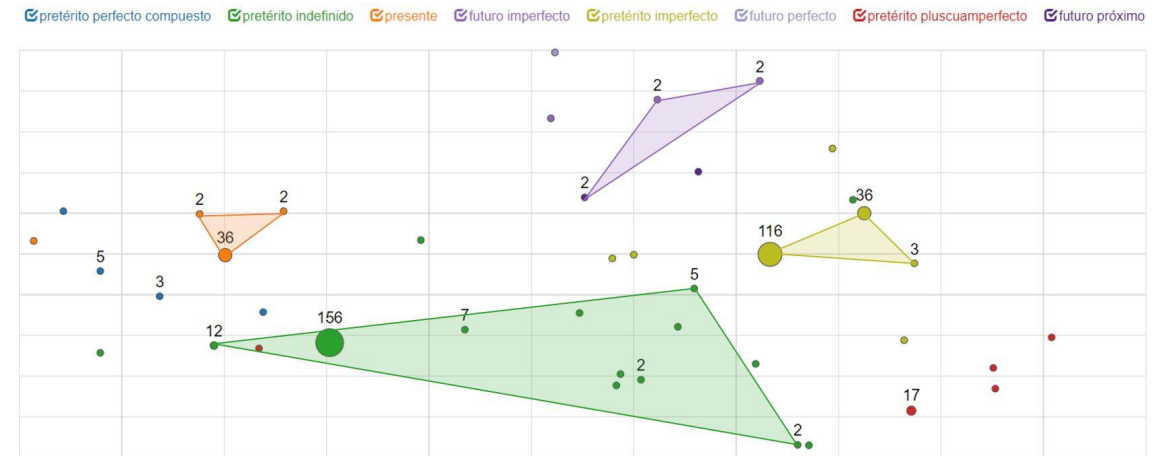
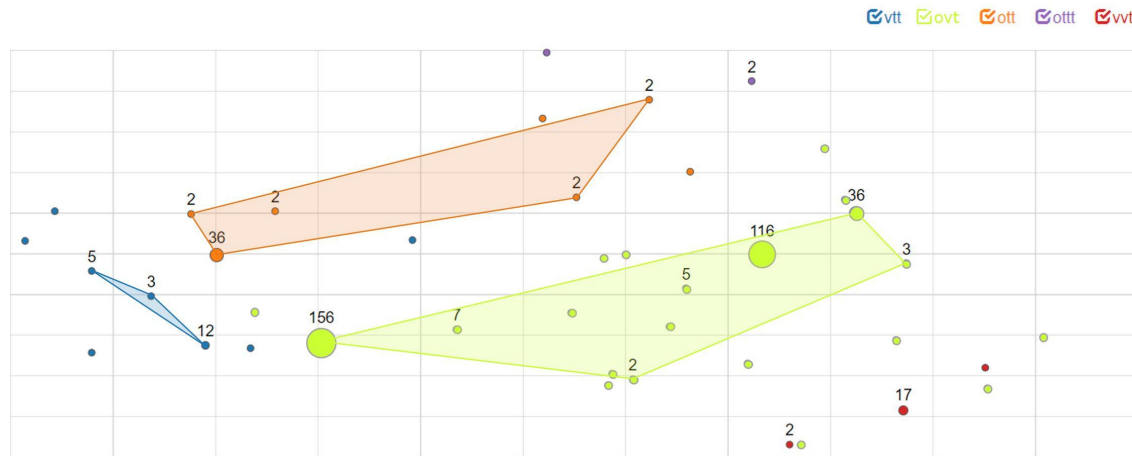
Perfekt
 Präteritum
 Präsens
 Futur I
 Futur II
 Plusquamperfekt



Translation Mining

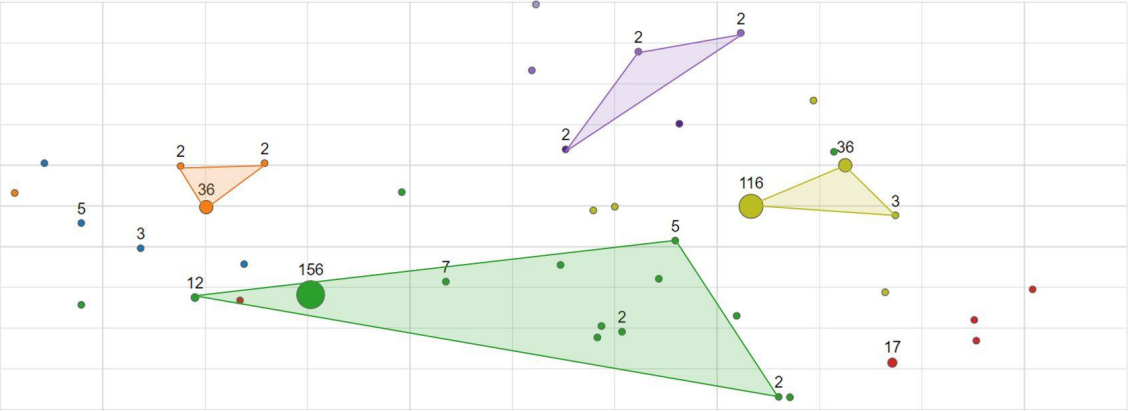


Translation Mining

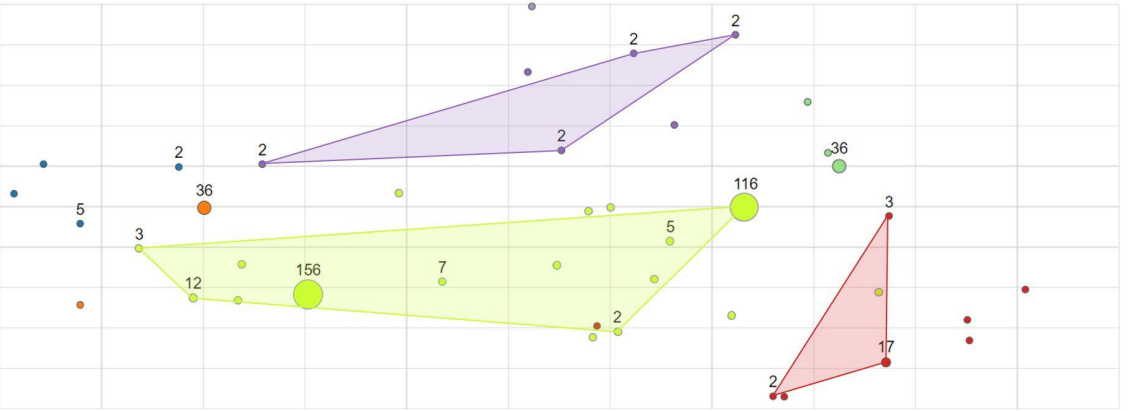


Translation Mining

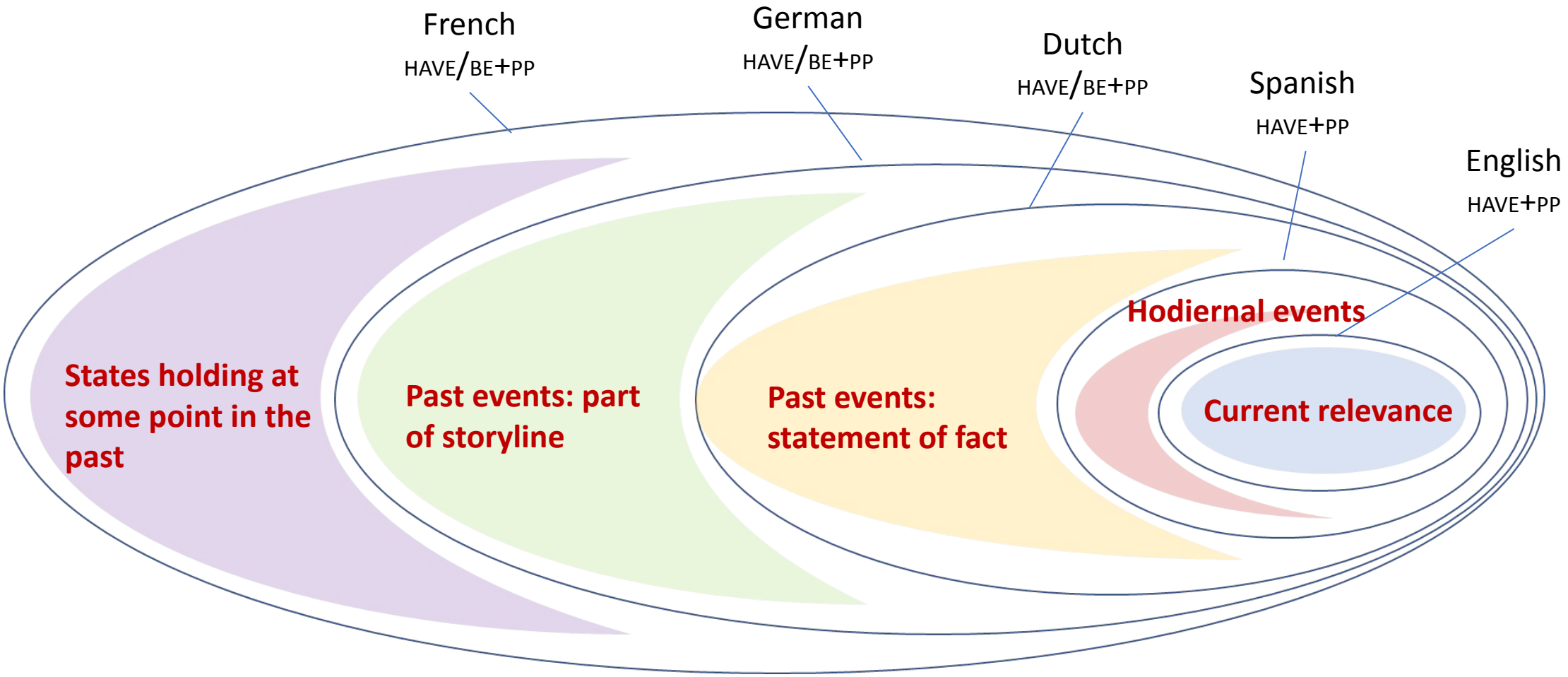
pretérito perfecto compuesto
 pretérito indefinido
 presente
 futuro imperfecto
 pretérito imperfecto
 futuro perfecto
 pretérito pluscuamperfecto
 futuro próximo



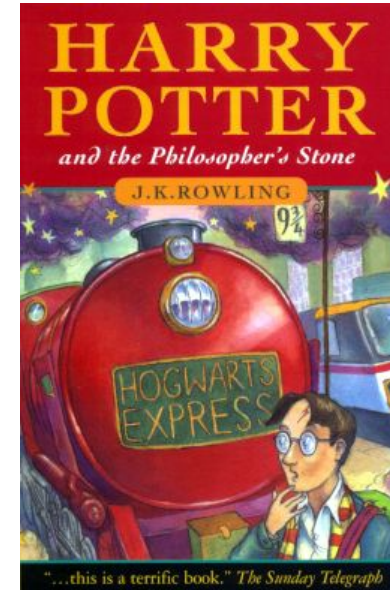
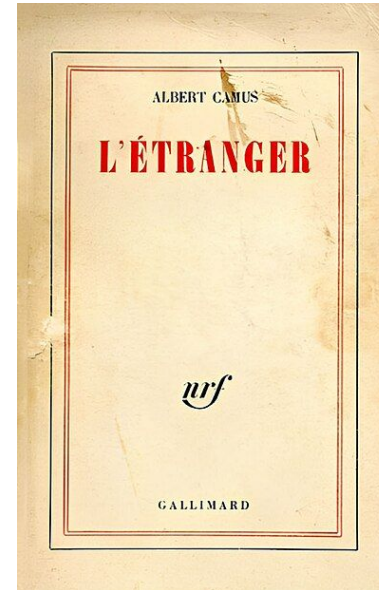
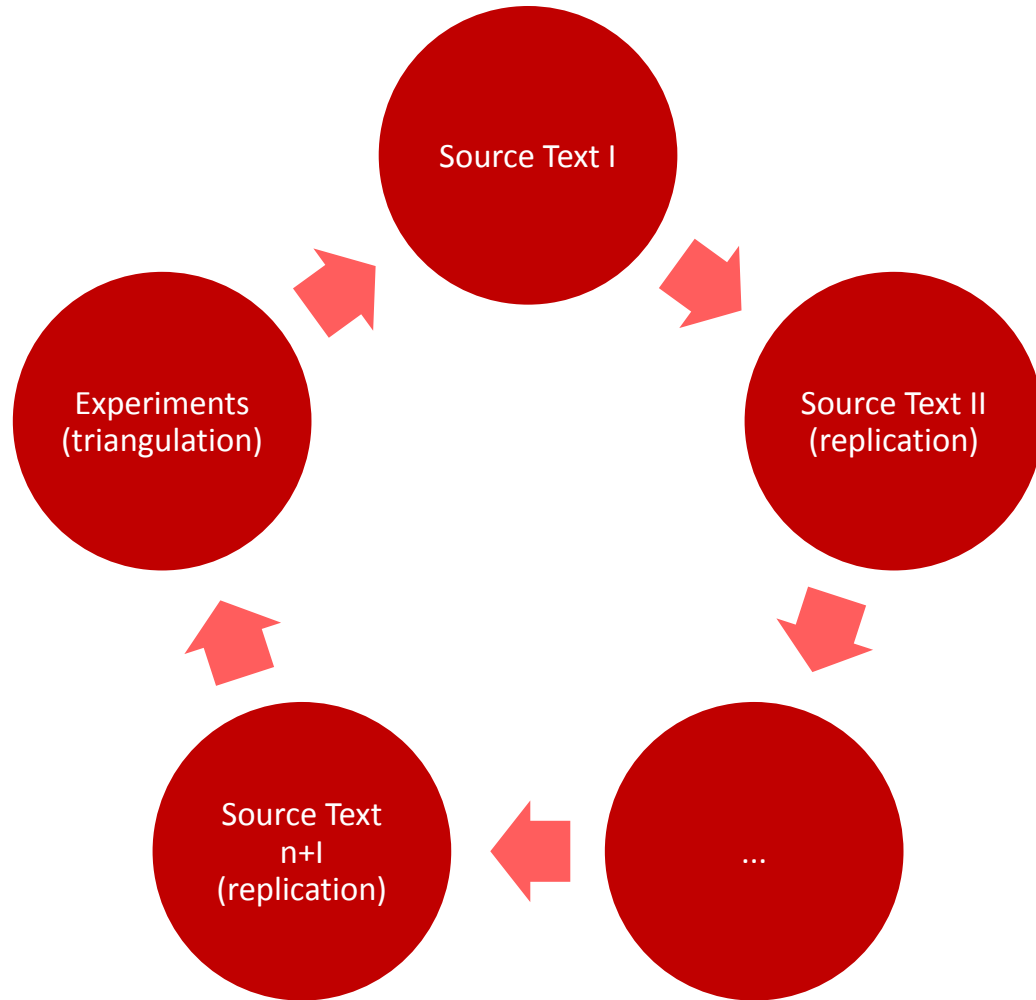
simple past
 present perfect
 simple present
 simple future
 future perfect
 past continuous
 past perfect



Translation Mining



Translation Mining



See Van der Klis et al. 2022 for the original *L'Étranger* study and Le Bruyn & de Swart (2022, 2023) for the results of the replication study. First experimental results are published in Fuchs & van der Klis (2023)

Beyond *Time in Translation*

Beyond *Time in*

Translation

> Original focus on **meso-variation** in the use of **HAVE/BE + past participle** in Western European languages.

> Move to **macro-variation**:

- **Chou Mo**'s dissertation work (Mo 2022): study of Mandarin **aspect** through the Mandarin translation of *L'Étranger*.
- **Jianan Liu**'s dissertation work (ongoing): study of **reference** in Mandarin, including cross-linguistic joint work with Olga Borik, Shravani Patil, Daria Seres, Hagay Schurr, and me *aka* the **HHRM group**.



Beyond *Time in*

Translation

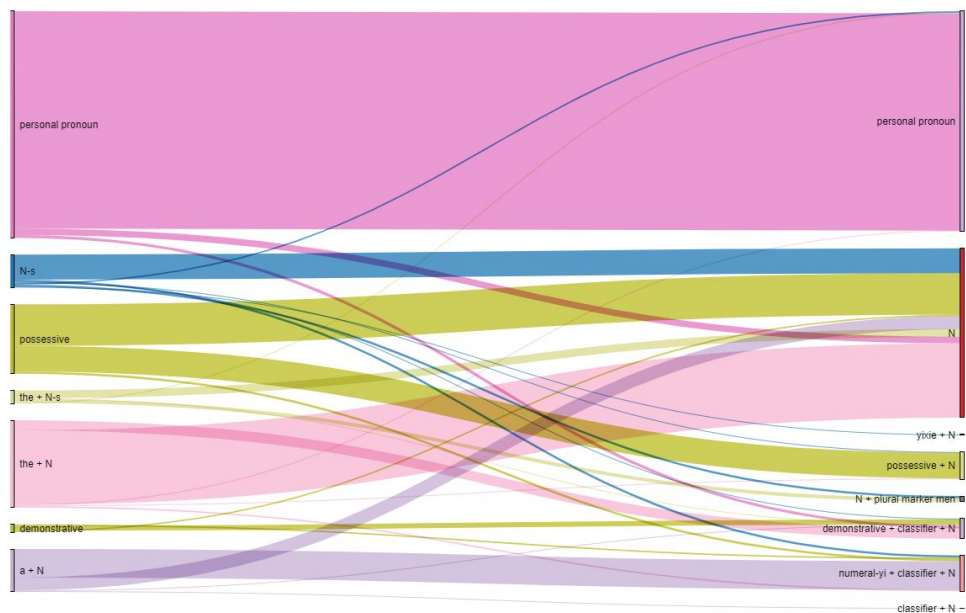
- > Reports on the work of the **HHRM group**, with a focus on **Mandarin**.
- > Argues that:
 - For **indefiniteness**: Mandarin has a full-fledged indefinite article and that bare nouns with an indefinite interpretation are (pseudo-)incorporated;
 - For **definiteness**: There are two types of strong definiteness, one relying on bare nouns, the other relying on the demonstrative;
 - **Methodology-wise**: *Translation Mining* can insightfully be applied to study macro-variation.

Basic data

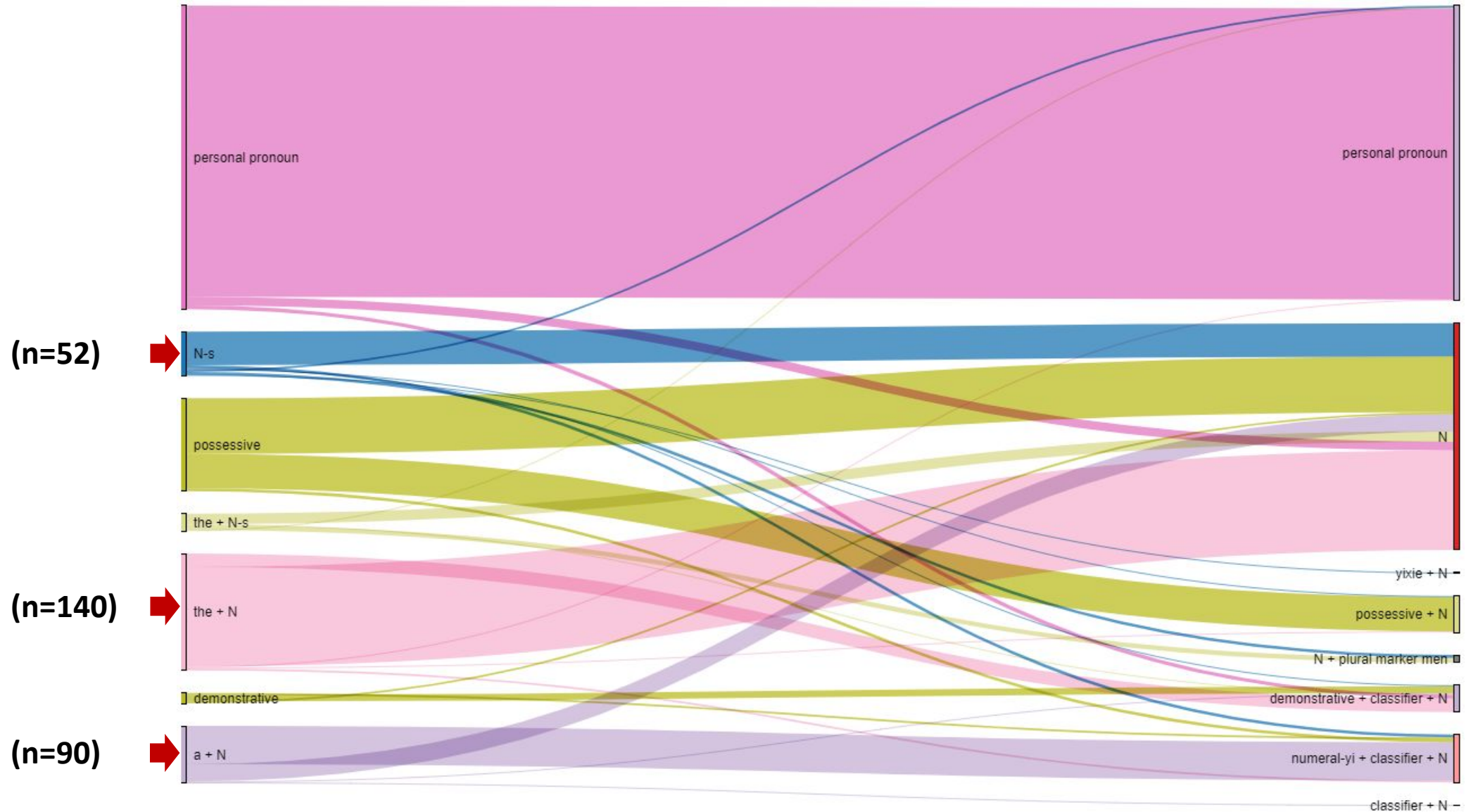
Basic data

Chapter 1, *Harry Potter and the Philosopher's Stone*

- > Translations to **Mandarin**.
- > Manual (computer-assisted) extraction of all nominal expressions (including pronouns) from the English original (N=1210).

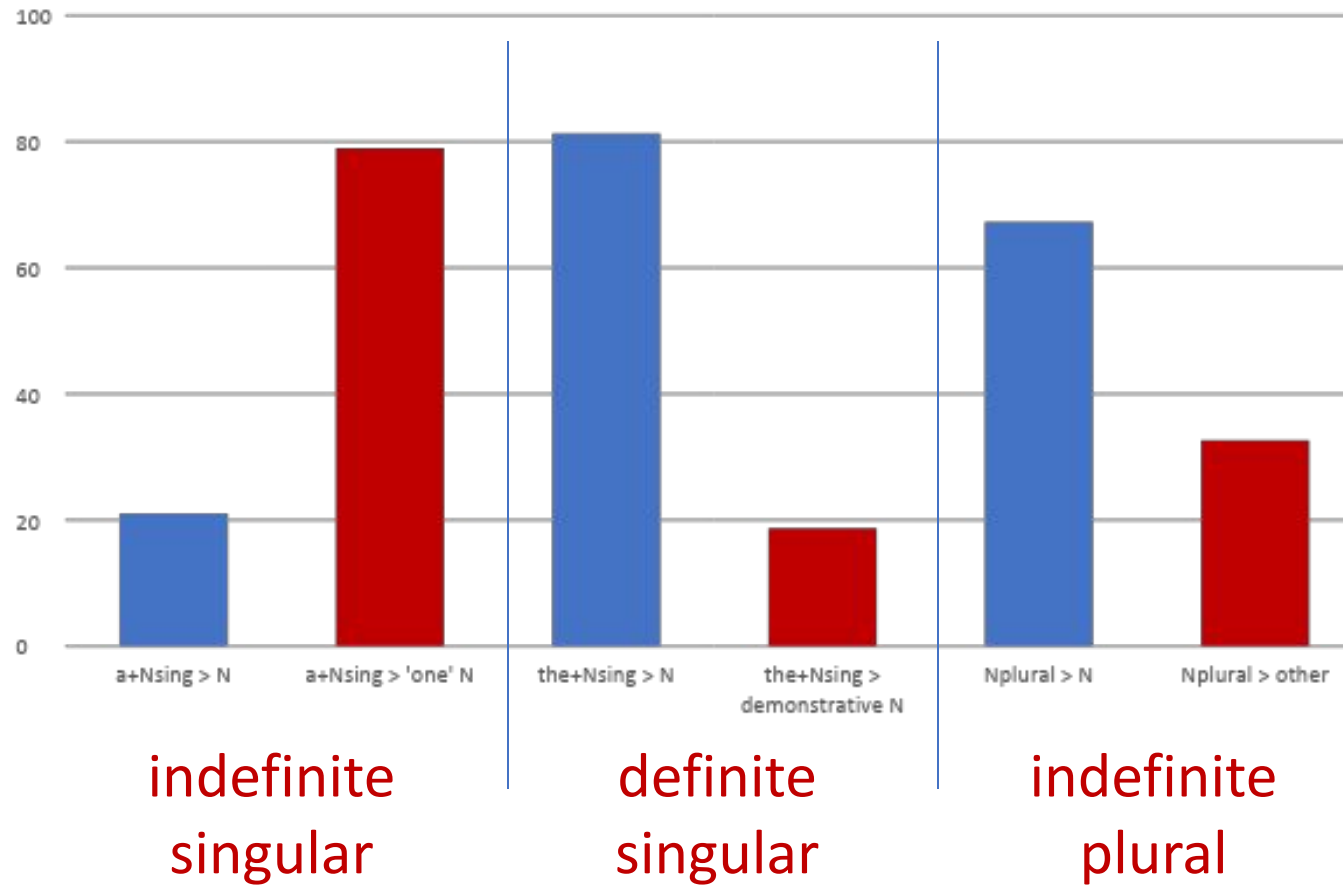


Basic data (selection)



Basic data (selection)

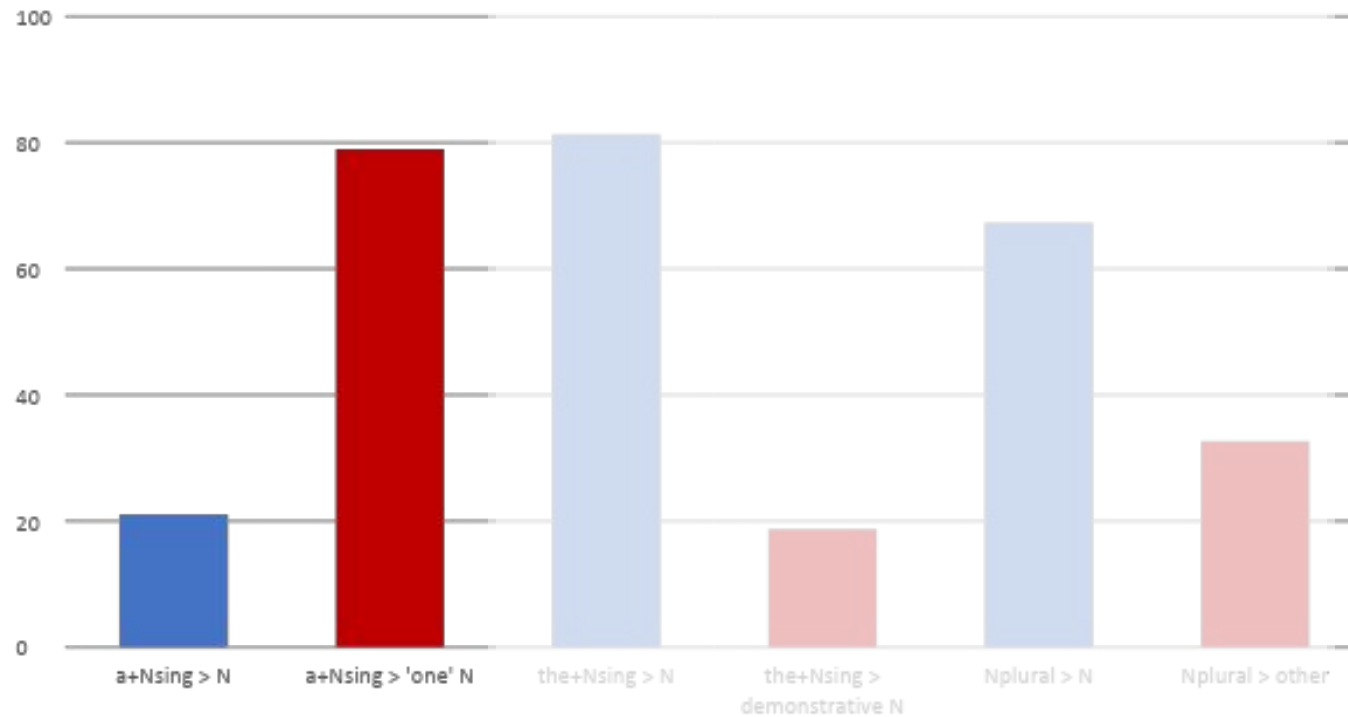
Relative proportions of N/DN translations
from English to Mandarin



Indefiniteness in Mandarin

Basic data

Relative proportions of N/DN translations
from English to Mandarin



indefinite
singular

Basic data

It was on the corner of the street that he noticed the first sign of something peculiar

- a cat reading a map

一只猫在看地图。

yī zhǐ māo zài kàn dìtú.

ONE CL cat ASP see map

Dumbledore gave a great sniff as he took a golden watch from his pocket [...]

从衣袋里掏出一块金表

cóng yīdài lǐ tāo chū yī kuài jīn biǎo

from pocket in pull out one CL gold watch

Basic data

It was on the corner of the street that he noticed the first sign of something peculiar

- a cat reading a map

一只猫在看地图。

yī zhǐ māo zài kàn dìtú.

ONE CL cat ASP see map

Dumbledore gave a great sniff as he took a golden watch from his pocket [...]

从衣袋里掏出一块金表

cóng yīdài lǐ tāo chū yī kuài jīn biǎo

from pocket in pull out one cl gold watch

Basic data

It was on the corner of the street that he noticed the first sign of something peculiar

- a cat reading a map
一只猫在看地图。
yī zhī māo zài kàn dìtú.
ONE CL cat ASP see map

Dumbledore gave a great sniff as he took a golden watch from his pocket [...]

从衣袋里掏出一块金表
cóng yīdài lǐ tāo chū yī kuài jīn biǎo
from pocket in pull out one CL gold watch

The formal literature – to the best of my knowledge – does not predict variation:

> Chierchia (1998)

> Dayal (2004)

> Cheng & Sybesma (1999)

> Jiang (2020)

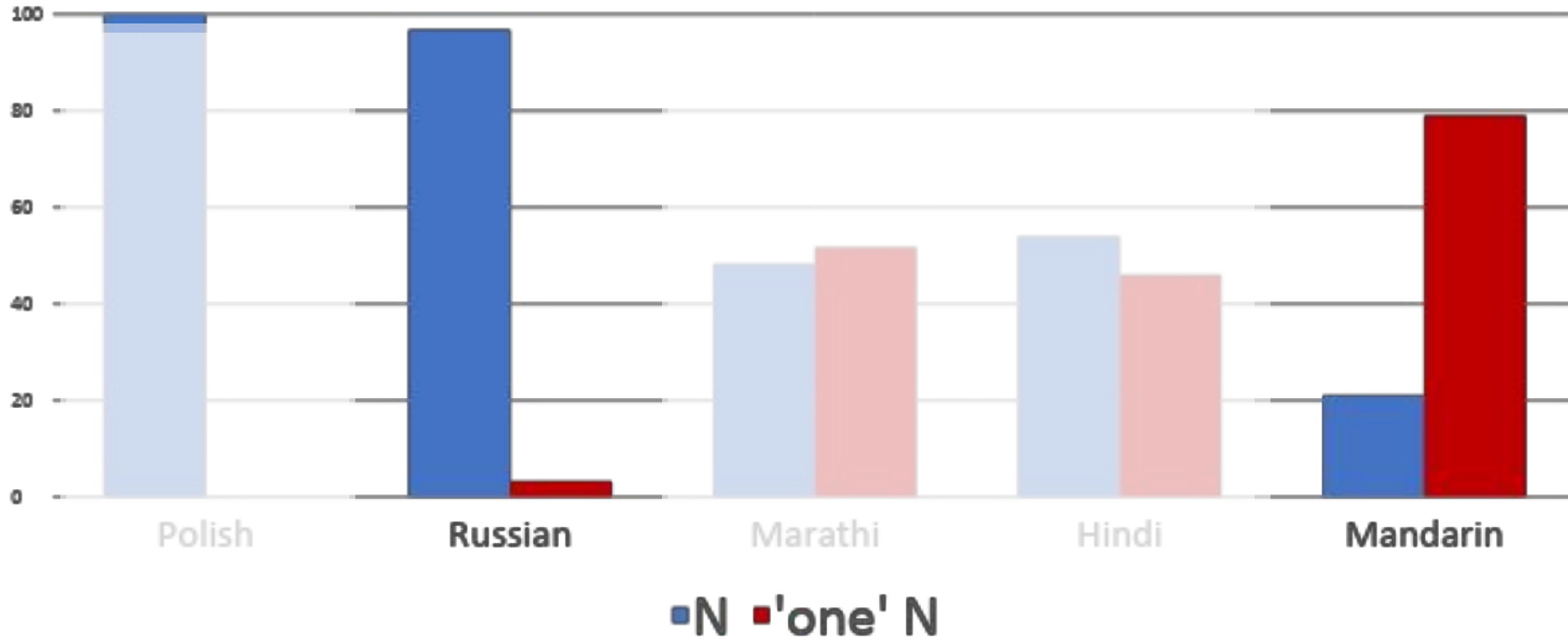
random variation?

Going cross-linguistic

Relative proportions of BN/'one' N translations from
English to Polish, Russian, Marathi, Hindi and
Mandarin

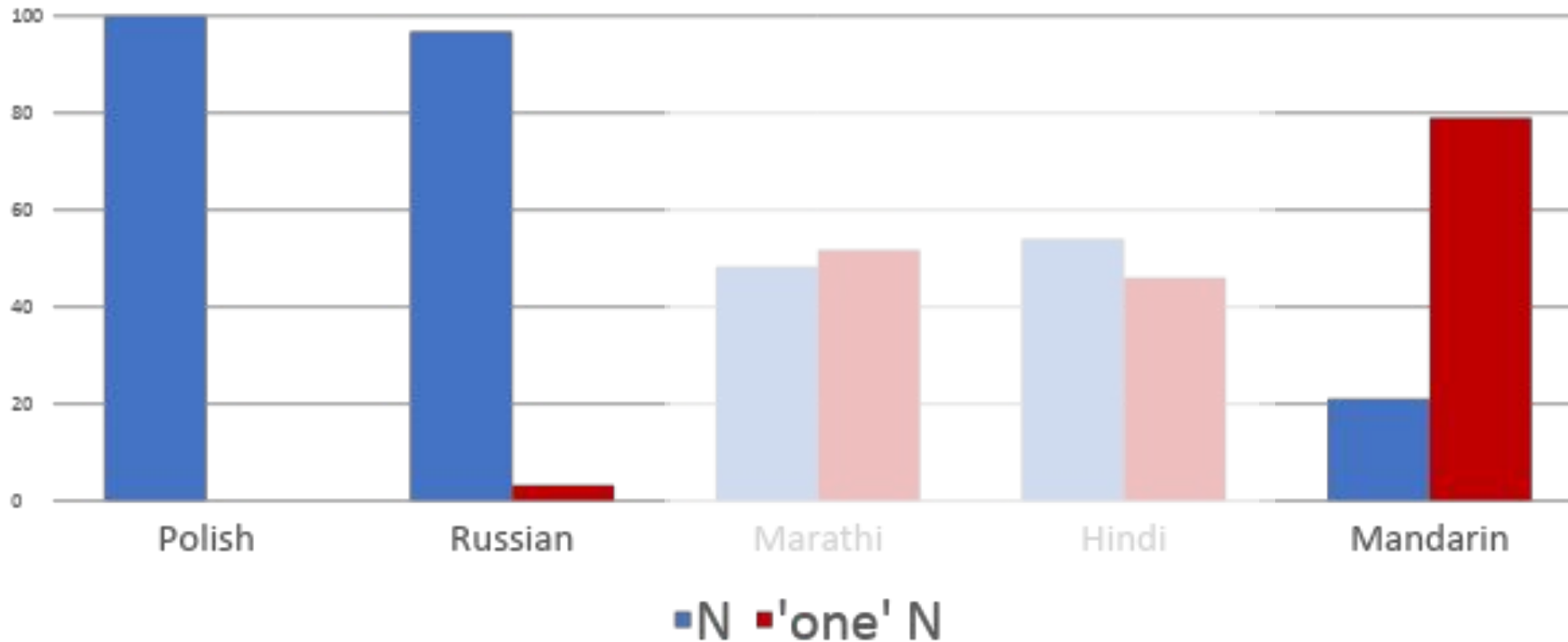
Going cross-linguistic

Relative proportions of BN/'one' N translations from English to Polish, Russian, Marathi, Hindi and Mandarin



Going cross-linguistic

Relative proportions of BN/'one' N translations from English to Polish, Russian, Marathi, Hindi and Mandarin



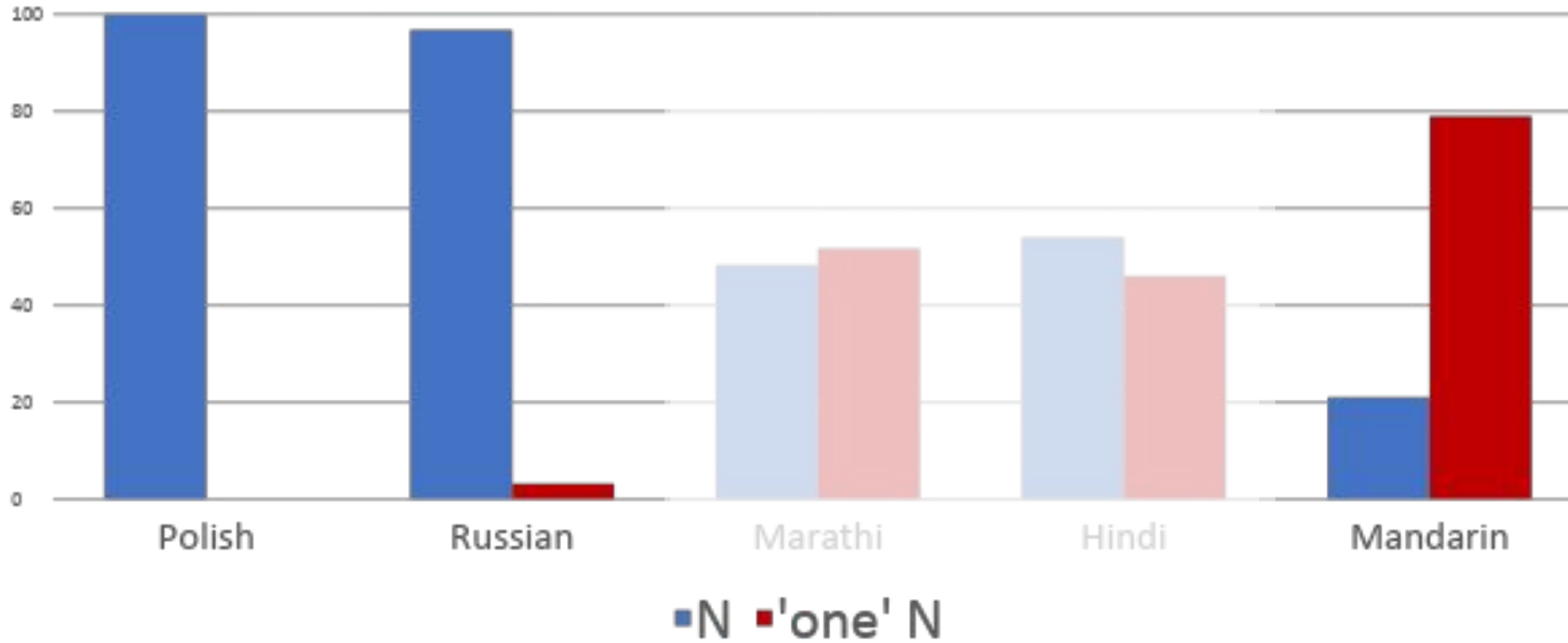
Going cross-linguistic

Interim conclusion

> The numeral 'one' uses in Mandarin **cannot be attributed to random variation**. If they could, we would expect similar 'random' uses in other articleless languages.

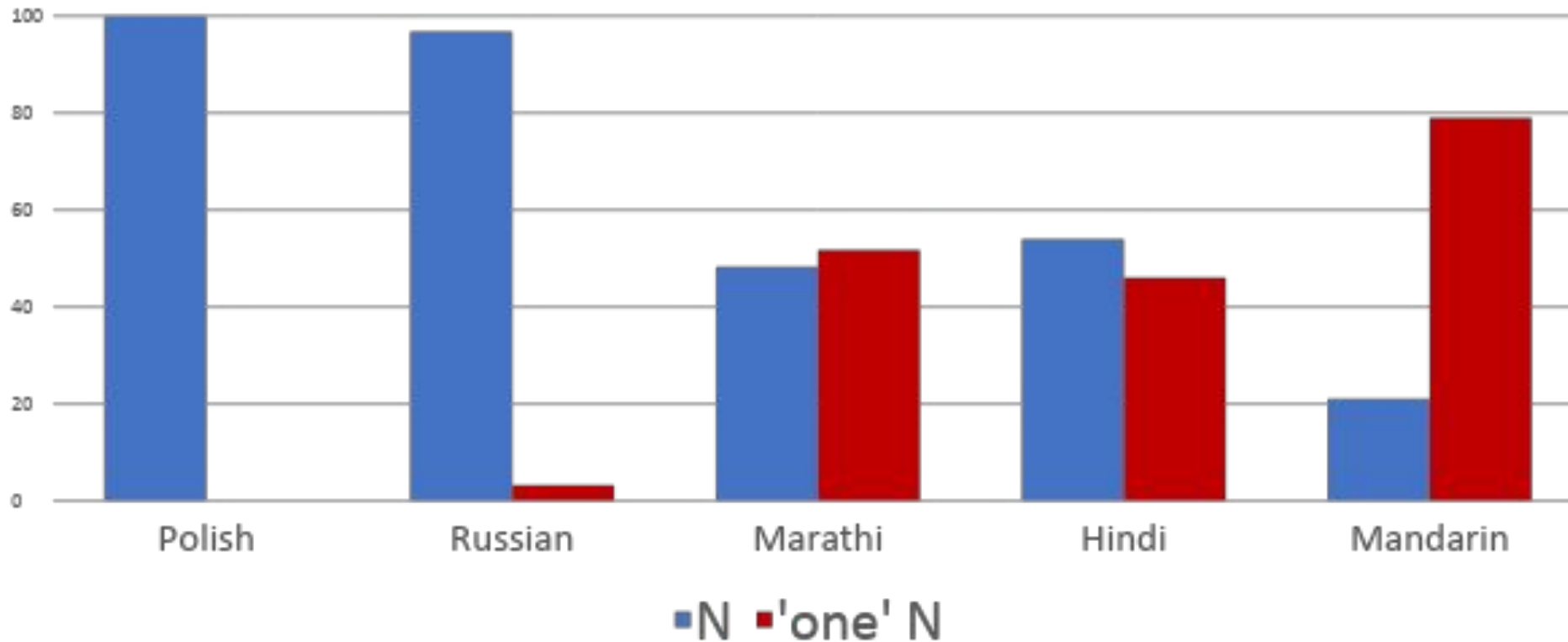
Going cross-linguistic

Relative proportions of BN/'one' N translations from English to Polish, Russian, Marathi, Hindi and Mandarin



Going cross-linguistic

Relative proportions of BN/'one' N translations from English to Polish, Russian, Marathi, Hindi and Mandarin



Going cross-linguistic (selection)

source id	Russian	Polish	Marathi	Hindi	Mandarin
152793	N	N	N	N	N
153248	N	N	N	N	N
153337	N	N	N	N	N
153721	N	N	N	N	N
153768	N	N	N	N	N
152853	N	N	N	N	'one' N
152867	N	N	N	N	'one' N
152895	N	N	N	N	'one' N
152955	N	N	N	N	'one' N
153589	N	N	N	N	'one' N
153769	N	N	N	N	'one' N
152880	N	N	N	'one' N	'one' N
153682	N	N	N	'one' N	'one' N
153317	N	N	'one' N	N	'one' N
152730	N	N	'one' N	'one' N	'one' N
153211	N	N	'one' N	'one' N	'one' N
153249	N	N	'one' N	'one' N	'one' N
153288	N	N	'one' N	'one' N	'one' N
153309	N	N	'one' N	'one' N	'one' N
153476	N	N	'one' N	'one' N	'one' N
153537	N	N	'one' N	'one' N	'one' N
153648	N	N	'one' N	'one' N	'one' N
153669	N	N	'one' N	'one' N	'one' N
153707	N	N	'one' N	'one' N	'one' N
153725	N	N	'one' N	'one' N	'one' N

Going cross-linguistic

Interim conclusion

- > The numeral 'one' uses in Mandarin **cannot be attributed to random variation**. If they could, we would expect similar random uses in other articleless languages like Russian.
- > The numeral 'one' uses in Hindi/Marathi and Mandarin seem to form **a scale**, suggesting that at least some of the factors influencing the choice between bare nouns and 'one' N are constant.

Dayal (2004, 2011), Huang (2015), Luo (2022)

> Dayal (2004, 2011):

- Hindi has a general ban on singular indefinite bare nouns requiring the use of an indefinite determiner (like, e.g., the numeral ‘one’).
- Bare nouns that do occur in singular indefinite environments are (pseudo-)incorporated.

> Huang (2015) suggests that (pseudo-)incorporation might also exist in Mandarin and manifests itself – among others – in ‘well-established’ verb-noun combinations: *kan shu* (‘read book’), *kai wanxiao* (‘make joke’).

> Luo (2022) is the first to propose a formal semantic analysis of pseudo-incorporation in Mandarin and builds on Schwarz (2014) but implements it in a kind approach to bare nouns.

Dayal (2004, 2011), Huang (2015), Luo (2022)

> Dayal (2004, 2011):

- Hindi has a general ban on singular indefinite bare nouns requiring the use of an indefinite determiner (like, e.g., the numeral ‘one’).
- Bare nouns that do occur in singular indefinite environments are (pseudo-)incorporated.

> Huang (2015) suggests that (pseudo-)incorporation might also exist in Mandarin and manifests itself – among others – in ‘well-established’ verb-noun combinations: *kan shu* (‘read book’), *kai wanxiao* (‘make joke’).

> Luo (2022) is the first to propose a formal semantic analysis of pseudo-incorporation in Mandarin and builds on Schwarz (2014) but implements it in a kind approach to bare nouns.

Dayal (2004, 2011), Huang (2015), Luo (2022)

> Dayal (2004, 2011):

- Hindi has a general ban on singular indefinite bare nouns requiring the use of an indefinite determiner (like, e.g., the numeral 'one').
- Bare nouns that do occur in singular indefinite environments are (pseudo-)incorporated.

> Huang (2015) suggests that (pseudo-)incorporation might also exist in Mandarin and manifests itself – among others – in 'well-established' verb-noun combinations: *kan shu* ('read book'), *kai wanxiao* ('make joke').

> Luo (2022) is the first to propose a formal semantic analysis of pseudo-incorporation in Mandarin and builds on Schwarz (2014) but implements it in a kind approach to bare nouns.

Hypothesis

The numeral 'one' is obligatory in Mandarin and bare nouns can only appear in pseudo-incorporation contexts.

[Further differences between Hindi/Marathi and Mandarin should be attributed to independent factors.]

Challenge

Unlike Hindi (and other (pseudo-)incorporation languages like Hungarian), Mandarin has no form-based criteria to decide whether we're dealing with (pseudo-)incorporation: it has rigid word order, no case markers and no generalized singular/plural distinction.

Dayal (2004, 2011), Huang (2015), Luo (2022)

> Dayal (2004, 2011):

- Hindi has a general ban on singular indefinite bare nouns requiring the use of an indefinite determiner (like, e.g., the numeral 'one').
- Bare nouns that do occur in singular indefinite environments are (pseudo-)incorporated.

> Huang (2015) suggests that (pseudo-)incorporation might also exist in Mandarin and manifests itself – among others – in 'well-established' verb-noun combinations: *kan shu* ('read book'), *kai wanxiao* ('make joke').

> Luo (2022) is the first to propose a formal semantic analysis of pseudo-incorporation in Mandarin and builds on Schwarz (2014) but implements it in a kind approach to bare nouns.

Hypothesis

The numeral 'one' is obligatory in Mandarin and bare nouns can only appear in pseudo-incorporation contexts.

[Further differences between Hindi/Marathi and Mandarin should be attributed to independent factors.]

Challenge

Unlike Hindi (and other (pseudo-)incorporation languages like Hungarian), Mandarin has no form-based criteria to decide whether we're dealing with (pseudo-)incorporation: it has rigid word order, no case markers and no generalized singular/plural distinction.

Dayal (2004, 2011), Huang (2015), Luo (2022)

> Dayal (2004, 2011):

- Hindi has a general ban on singular indefinite bare nouns requiring the use of an indefinite determiner (like, e.g., the numeral 'one').
- Bare nouns that do occur in singular indefinite environments are (pseudo-)incorporated.

> Huang (2015) suggests that (pseudo-)incorporation might also exist in Mandarin and manifests itself – among others – in 'well-established' verb-noun combinations: *kān shū* ('read book'), *kāi wánxiào* ('make joke').

> Luo (2022) is the first to propose a formal semantic analysis of pseudo-incorporation in Mandarin and builds on Schwarz (2014) but implements it in a kind approach to bare nouns.

Hypothesis

The numeral 'one' is obligatory in Mandarin and bare nouns can only appear in pseudo-incorporation contexts.

[Further differences between Hindi/Marathi and Mandarin should be attributed to independent factors.]

Challenge

Unlike Hindi (and other (pseudo-)incorporation languages like Hungarian), Mandarin has no form-based criteria to decide whether we're dealing with (pseudo-)incorporation: it has rigid word order, no case markers and no generalized singular/plural distinction.

Rising to the challenge

For Dayal, Huang and Luo, a crucial property of (pseudo-)incorporation is that – as far as the object position is concerned – it is limited to 'well-established' verb-noun combinations.

If we can operationalize what it means to be a 'well-established' verb-noun combination, we can check whether the bare nouns we find in object position in singular indefinite contexts in Mandarin are part of this type of combinations, allowing for a meaning-based assessment of their (pseudo-)incorporation status.

Le Bruyn et al. (2016) and Liu et al. (2022)

- > Le Bruyn et al. (2016) is cast in dynamic predicate logic and analyzes pseudo-incorporation of verb-noun combinations as involving a verb that targets an explicit or implicit argument of its object noun and prepares it for binding by its subject. Illustrative examples are **HAVE SON** (explicit relational argument) and **READ BOOK** (implicit qualia-based 'user' argument).
- > In Liu et al. (2022), we argue that the way Le Bruyn et al. (2016) model pseudo-incorporation derives the well-establishedness constraint rather than stipulating it.
- > In this talk, I use it as a way to decide whether verb-noun combinations qualify as allowing for pseudo-incorporation.

An extended (and more focused) dataset

- > The basic data I have used up till now is limited to the first chapter of *Harry Potter and the Philosopher's Stone* and includes translations of *a* *N* in all possible positions.
- > Our operationalization of our (pseudo-)incorporation check is limited to nouns in object position.
- > We keep the object noun data from the first chapter but extend them with object noun data from the whole book (N=156).

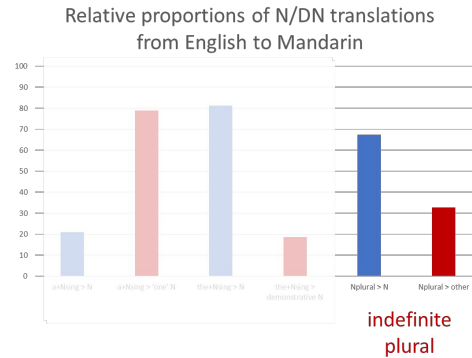
Annotation and analysis

- > We annotated all the bare noun and 'one' N indefinite objects in their Mandarin translations for:
 - `form`: is the indefinite object rendered as a bare noun (N) or as a structure with a numeral ('one' N)?
 - `incorporation`: does the V-O combination it occurs in qualify as 'well-established'?
- > We rely on **conditional inference trees** (Tagliamonte & Baayen 2012) to assess whether `incorporation` is a good predictor for `form`.

Further annotations

- > We further annotated all datapoints for a number of other characteristics that we assumed could influence `form`:
 - `modification`: is the noun modified or not? (yes/no)
 - `non-referential`: does the noun refer to an actual individual in the universe of the story? (yes/no)
 - `aspect`: does the aspect marker `le` occur on the verb? (yes/no)
 - `NEG/DIS`: see next slide (yes/no)

Further annotations: NEG/DIS



Most of them had never seen **an owl** even at nighttime .

他们大多甚至夜里都从未见过猫头鹰。

‘ You must come and stay this summer , ’ said Ron , ‘ both of you - I 'll send you **an owl**. ’

“你今年暑假一定要来我们家里玩，”罗恩说，“你们俩都来——我会派*猫头鹰*去邀请你们的。”

Some examples

It was on the corner of the street that he noticed the first sign of something peculiar

- a cat reading a map

一只猫在看地图。

yī zhǐ māo zài kàn dìtú.

ONE CL cat ASP see map

Dumbledore gave a great sniff as he took a golden watch from his pocket [...]

从衣袋里掏出一块金表

cóng yīdài lǐ tāo chū yī kuài jīn biǎo

from pocket in pull out one CL gold watch

Some examples

A ghost wearing a ruff and tights had suddenly noticed the first-years .

一个穿轮状皱领紧身衣的幽灵突然发现了一年级新生。

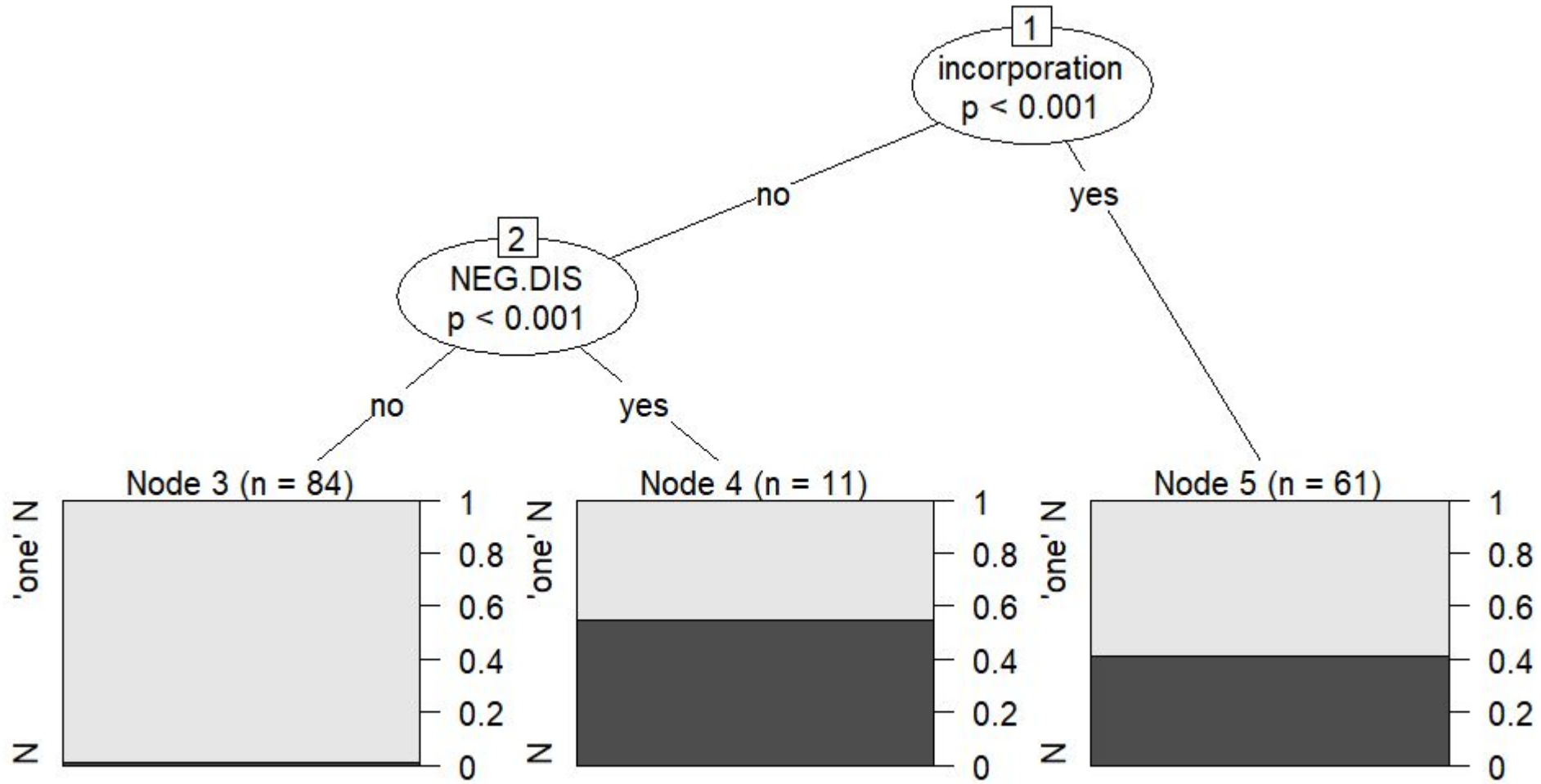
A barn owl brought Neville a small package from his grandmother .

一只猫头鹰从纳威的奶奶那里给他带来了一个小包裹。

Quirrell raised his hand to perform a deadly curse, but Harry, by instinct, reached up and grabbed Quirrell 's face -

奇洛举起手，准备念一个死咒，可是哈利出于本能，猛地抬手抓向奇洛的脸——

Results



The exception

It was on the corner of the street that he noticed the first sign of something peculiar - a cat reading a map. For a second, Mr Dursley didn't realise what he had seen - then he jerked his head around to look again. There was a tabby cat standing on the corner of Privet Drive, but there wasn't **a map** in sight.

但是 没有 看见 地图。

dànshì méiyǒu kànjiàn **dìtú** .

but not see **map**

Talk

overview

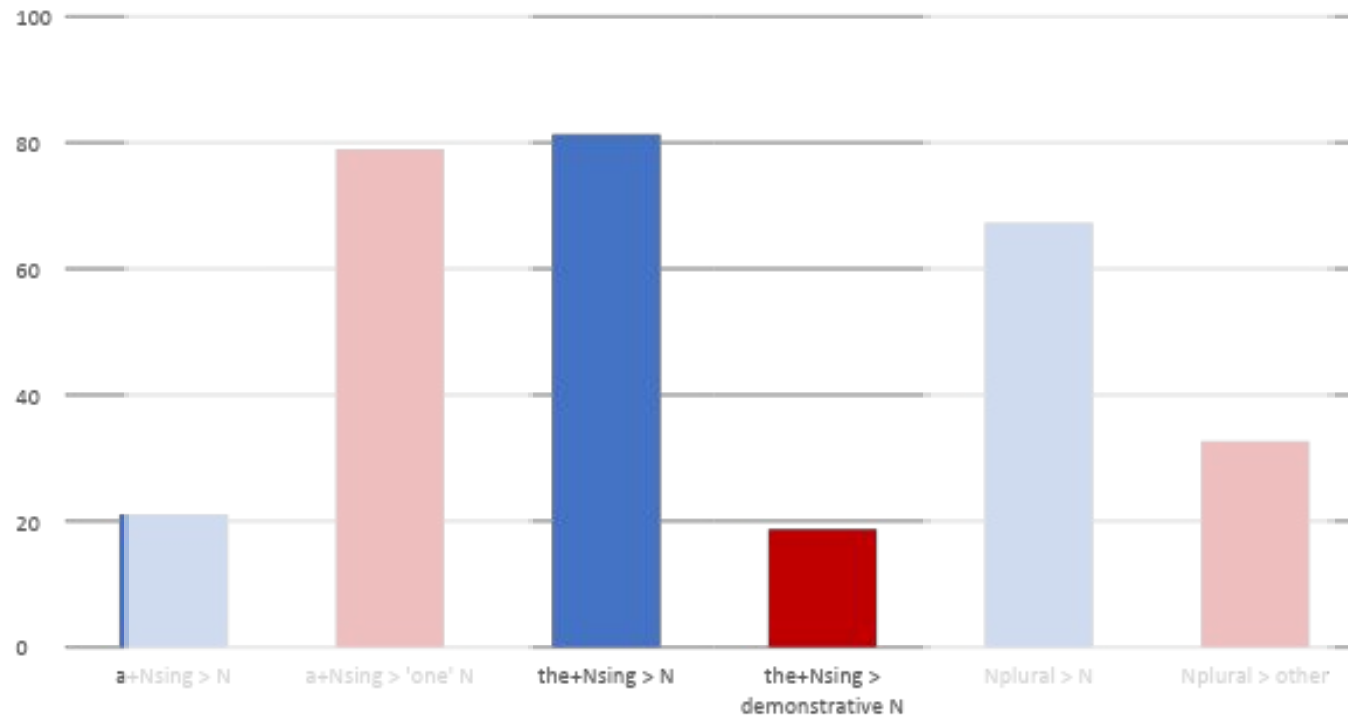
- > Reports on the work of the **HHRM group**, with a focus on **Mandarin**.
- > Argues that:
 - For **indefiniteness**: Mandarin has a full-fledged indefinite article and that bare nouns with an indefinite interpretation are (pseudo-)incorporated;
 - For **definiteness**: There are two types of strong definiteness, one relying on bare nouns, the other relying on the demonstrative;
 - **Methodology-wise**: *Translation Mining* can insightfully be applied to study macro-variation.



Definiteness in Mandarin

Basic data

Relative proportions of N/DN translations
from English to Mandarin



definite
singular

Going cross-linguistic I

> Intuition from Jenks (2018): the contrast we find between bare nouns and demonstratives in Mandarin is the same as the one we find between weak and strong definites in the German prepositional domain.

Going cross-linguistic I

(#Nà /#zhè ge) táiwān (de) zǒngtǒng hěn shēngqì.

that /this CLF Taiwan('s) president very angry

‘The president of Taiwan is very angry.’ (Jenks 2018:507)

Jiàoshì lǐ zuò-zhe yī gè nánshēng hé yī gè
classroom in sit-ASP one CLF boy and one classifier

nǚshēng. Wǒ zuótiān yùdào #(nà gè) nánshēng
girl I yesterday meet that CLF boy.

‘There are a boy and a girl sitting in the classroom. I met **the boy** yesterday.’

(Jenks 2018:510)

Going cross-linguistic I

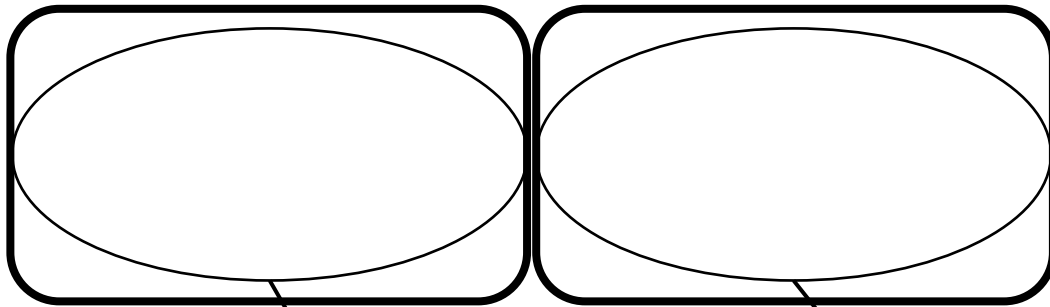
- > Intuition from Jenks (2018): the contrast we find between bare nouns and demonstratives in Mandarin is the same as the one we find between weak and strong definites in the German prepositional domain.
- > Bremmers et al. (2022): contrastive study based on *Harry Potter and the Philosopher's Stone* (full volume) where we compare German PPs with weak and strong definites to bare nouns and demonstratives in Mandarin.

Going cross-linguistic I

Prediction

German uncontr. def.

German contr. def.



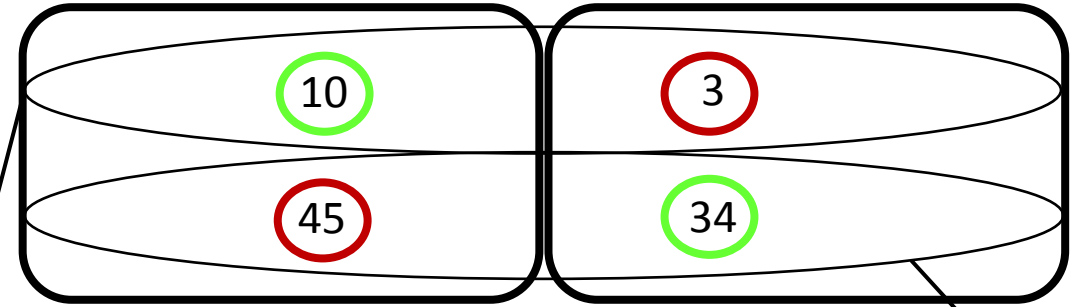
Mandarin dem.

Mandarin BN

Results

German uncontr. def.

German contr. def.



Mandarin dem.

Mandarin BN

Going cross-linguistic I

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

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

M: Pèinī, wǒ juébù ràng tāmen rènherén jìn zhè dòng fángzi.’

Petunia I not have them anyone enter this CLF house

Going cross-linguistic I

[Context: 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
they were hardly out the way fluttered when a other owl
einen Brief **auf das Paket** warf.
a letter on the parcel threw

M: Tāmen pūshan-zhe chìbǎng gānggāng fēi zǒu, yòu
They flutter-ASP wings right fly away, and
yǒu yī zhǐ māotóuyīng xié lái yī fēng xìn,
have one classifier owl bring come one CLF letter
rēng zài **bāoguǒ** shàngmiàn.
throw to parcel on.

Going cross-linguistic I

Interim conclusion

There are strong indications that Mandarin relies on bare nouns to convey weak definiteness and that demonstratives are sometimes used for strong definiteness. However, there is no one-to-one mapping between demonstratives and strong definiteness given that bare nouns also regularly appear in familiarity contexts.

Ahn (2019), Simpson & Wu (2022), and Dayal & Jiang (2022)

- > Researchers following up on Jenks (2018) have focused on the choice between bare nouns and demonstratives in familiarity contexts.
- > I'll take the example below as my running example in the discussion of three follow-up proposals:

Jiàoshì	lǐ	zuò-zhe	yī	gè	nánshēng	hé	yī	gè
classroom	in	sit-ASP	one	CLF	boy	and	one	classifier
nǚshēng.	Wǒ	zuótiān	yùdào	#(nà	gè)	nánshēng		
girl	I	yesterday	meet	that	CLF	boy.		

‘There are a boy and a girl sitting in the classroom. I met **the boy** yesterday.’

(Jenks 2018:510)

Ahn (2019), Simpson & Wu (2022), and Dayal & Jiang (2022)

> According to Ahn (2019), the unacceptability of the bare noun is not as strict as it is presented by Jenks (2018). She assumes there is full optionality and relates this to the interaction between two opposing competitions: a competition that favors the use of the bare noun over the demonstrative (*Don't overdetermine!*) and another competition that favors the use of the demonstrative over that of the bare noun (*Index!*).

Ahn (2019), Simpson & Wu (2022), and Dayal & Jiang (2022)

jiàoshì	lǐ	zuò-zhe	yī	gè	nánshēng	hé	yī	gè
classroom	in	sit-ASP	one	CLF	boy	and	one	classifier
nǚshēng.	Wǒ	zuótiān	yùdào	#(nà	gè)	nánshēng		
girl	I	yesterday	meet	that	CLF	boy.		

‘There are a boy and a girl sitting in the classroom. I met **the boy** yesterday.’

(Jenks 2018:510)

Ahn (2019), Simpson & Wu (2022), and Dayal & Jiang (2022)

- > According to Simpson & Wu (2022), bare nouns can behave like strong definites but an overt marker like a demonstrative might end up being preferred if the sentence that the strong definite appears in is less 'tightly connected' to the sentence its antecedent appears in.

Ahn (2019), Simpson & Wu (2022), and Dayal & Jiang (2022)

Jiàoshì lǐ zuò-zhe yī gè nánshēng hé yī gè
classroom in sit-ASP one CLF boy and one classifier
nǚshēng. Wǒ zuótiān yùdào #(nà gè) nánshēng
girl I yesterday meet that CLF boy.

‘There are a boy and a girl sitting in the classroom. I met **the boy** yesterday.’

[Context: 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.]

M: Tāmen pūshan-zhe chibǎng gānggāng fēi zǒu, yòu
They flutter-ASP wings right fly away, and
yǒu yī zhǐ māotóuyīng xié lái yī fēng xìn,
have one classifier owl bring come one CLF letter
rēng zài bāoguǒ shàngmiàn.
throw to parcel on.

Ahn (2019), Simpson & Wu (2022), and Dayal & Jiang (2022)

- > According to Simpson & Wu (2022), bare nouns can behave like strong definites but an overt marker like a demonstrative might end up being preferred if the sentence that the strong definite appears in is less ‘tightly connected’ to the sentence its antecedent appears in.
- > Simpson & Wu’s proposal depends on the assumption that **demonstratives are ‘stronger’ than strong definites.**

Ahn (2019), Simpson & Wu (2022), and Dayal & Jiang (2022)

- > Dayal & Jiang (2022) assume that bare nouns and demonstratives can be used in anaphoric contexts. For them, the crucial difference lies in that the bare noun is incompatible with non-unique referents whereas the **demonstrative requires non-uniqueness**.
- > According to them, ‘less tightly connected’ sentences (in Simpson and Wu’s terminology) indicate shifts between situations and **unless the speaker is absolutely sure that a situation shift has not made available any new referents, the demonstrative is considered to be the safest (optimal) bet.**

Ahn (2019), Simpson & Wu (2022), and Dayal & Jiang (2022)

Interim conclusion

- > The current consensus in the literature seems to be that there are pragmatic and/or semantic factors that influence the choice between bare nouns and demonstratives in familiarity contexts. There is consequently no optionality in the way proposed by Ahn (2019).
- > Simpson & Wu (2022) and Dayal & Jiang (2022) converge on the idea that sentences can be more or less tightly connected. Simpson & Wu capture this pragmatically whereas Dayal & Jiang spell this out semantically.

Ahn (2019), Simpson & Wu (2022), and Dayal & Jiang (2022)

Interim conclusion

> To relate the preference for demonstratives in 'less tightly connected' sentences, Simpson & Wu rely on a general accessibility scale (as discussed in Ahn 2019) whereas Dayal & Jiang rely on the anti-uniqueness nature of demonstratives.

Going cross-linguistic II

> To relate the preference for demonstratives in 'less tightly connected' sentences, Simpson & Wu rely on a general accessibility scale (as discussed in Ahn 2019) whereas Dayal & Jiang rely on the anti-uniqueness nature of demonstratives.



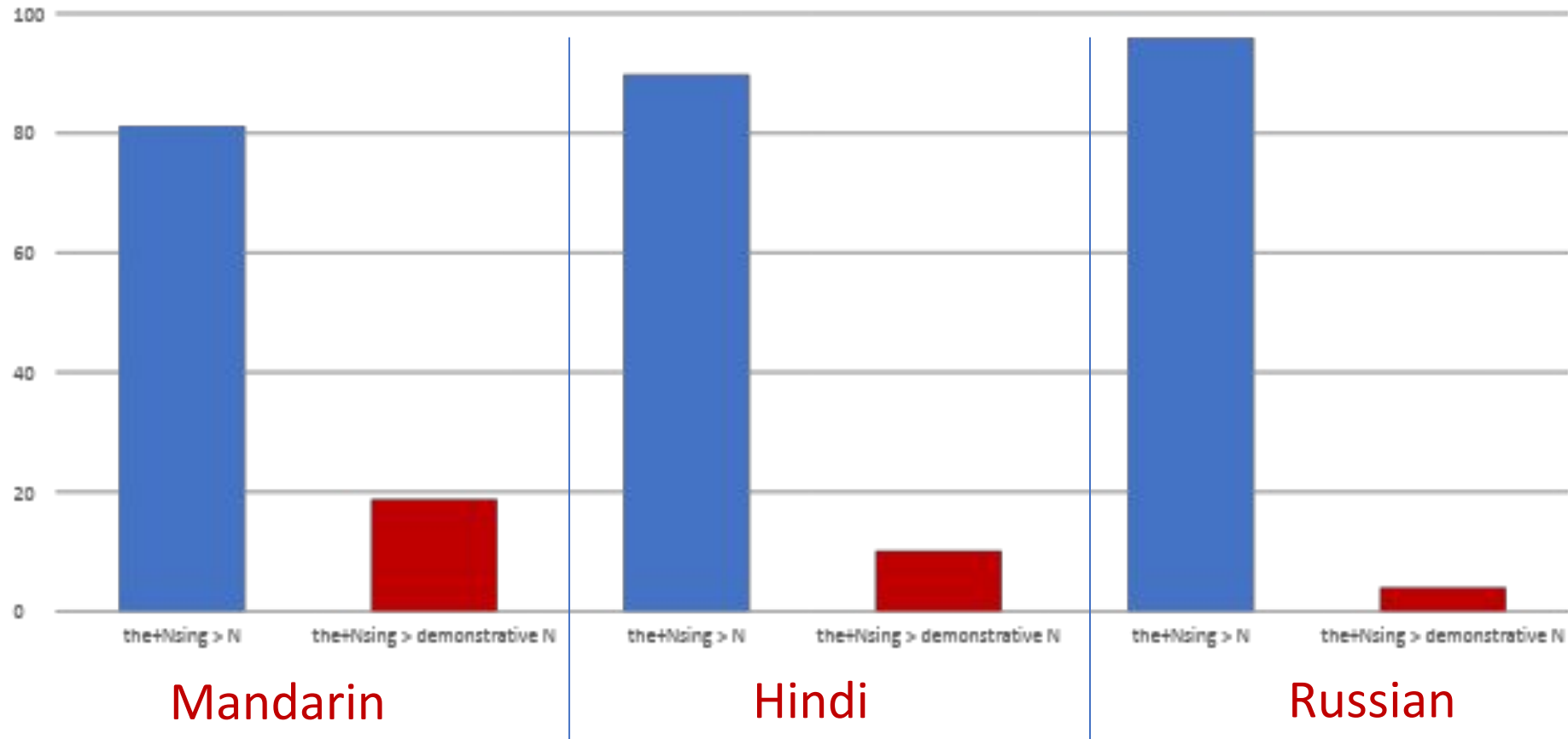
Both Simpson & Wu and Dayal & Jiang rely on **general** properties of demonstratives to explain the preference for a demonstrative in our running example. If correct, we would expect the same preference to pop up in other languages without definite articles.

Going cross-linguistic II

Relative proportions of BN/demonstrative N
translations from English to Mandarin, Hindi and
Russian

Going cross-linguistic II

Relative proportions of BN/demonstrative N translations from English to Mandarin, Hindi and Russian



17 anaphoric definites that take a demonstrative in at least one language:

14 in Mandarin
4 in Hindi
3 in Russian

Going cross-linguistic II

At half past eight, Mr Dursley picked up his briefcase, pecked Mrs Dursley on the cheek and tried to kiss Dudley goodbye but missed, because Dudley was now having a tantrum and throwing his cereal at the walls. 'Little tyke,' chortled Mr Dursley as he left the house. He got into his car and backed out of number four's drive.

It was on the corner of the street that he noticed the first sign of something peculiar - a cat reading a map. For a second, Mr Dursley didn't realise what he had seen - then he jerked his head around to look again. There was a tabby cat standing on the corner of Privet Drive, but there wasn't a map in sight. What could he have been thinking of? It must have been a trick of the light. Mr Dursley blinked and stared at the cat. It stared back.

As Mr Dursley drove around the corner and up the road, he watched the cat in his mirror. It was now reading the sign that said Privet Drive - no, looking at the sign; cats couldn't read maps or signs. Mr Dursley gave himself a little shake and put the cat out of his mind.

Going cross-linguistic II

At half past eight, Mr Dursley picked up his briefcase, pecked Mrs Dursley on the cheek and tried to kiss Dudley goodbye but missed, because Dudley was now having a tantrum and throwing his cereal at the walls. 'Little tyke,' chortled Mr Dursley as he left the house. **He got into his car and backed out of number four's drive.**

It was on the corner of the street that he noticed the first sign of something peculiar - a cat reading a map. For a second, Mr Dursley didn't realise what he had seen - then he jerked his head around to look again. There was a tabby cat standing on the corner of Privet Drive, but there wasn't a map in sight. What could he have been thinking of? It must have been a trick of the light. Mr Dursley blinked and stared at the cat. It stared back.

As Mr Dursley drove around the corner and up the road, he watched the cat in his mirror. It was now reading the sign that said Privet Drive - no, looking at the sign; cats couldn't read maps or signs. Mr Dursley gave himself a little shake and put the cat out of his mind.

Going cross-linguistic II

At half past eight, Mr Dursley picked up his briefcase, pecked Mrs Dursley on the cheek and tried to kiss Dudley goodbye but missed, because Dudley was now having a tantrum and throwing his cereal at the walls. 'Little tyke,' chortled Mr Dursley as he left the house. **He got into his car and backed out of number four's drive.**

It was on the corner of the street that he noticed the first sign of something peculiar - **a cat** reading a map. For a second, Mr Dursley didn't realise what he had seen - then he jerked his head around to look again. There was **a tabby cat** standing on the corner of Privet Drive, but there wasn't a map in sight. What could he have been thinking of? It must have been a trick of the light. Mr Dursley blinked and stared at the cat. It stared back.

As Mr Dursley drove around the corner and up the road, he watched the cat in his mirror. It was now reading the sign that said Privet Drive - no, looking at the sign; cats couldn't read maps or signs. Mr Dursley gave himself a little shake and put the cat out of his mind.

Going cross-linguistic II

At half past eight, Mr Dursley picked up his briefcase, pecked Mrs Dursley on the cheek and tried to kiss Dudley goodbye but missed, because Dudley was now having a tantrum and throwing his cereal at the walls. 'Little tyke,' chortled Mr Dursley as he left the house. **He got into his car and backed out of number four's drive.**

It was on the corner of the street that he noticed the first sign of something peculiar - **a cat** reading a map. For a second, Mr Dursley didn't realise what he had seen - then he jerked his head around to look again. There was **a tabby cat** standing on the corner of Privet Drive, but there wasn't a map in sight. What could he have been thinking of? It must have been a trick of the light. Mr Dursley blinked and stared at **the cat**. **It** stared back.

As Mr Dursley drove around the corner and up the road, he watched **the cat** in his mirror. **It** was now reading the sign that said Privet Drive - no, looking at the sign; cats couldn't read maps or signs. Mr Dursley gave himself a little shake and put **the cat** out of his mind.

Going cross-linguistic II

It was on the corner of the street [...] **a cat** [...] Mr Dursley blinked and stared at **the cat**. **It** stared back.

As Mr Dursley drove around the corner and up the road, he watched **the cat** in his mirror. **It** was now reading the sign that said Privet Drive [...] Mr Dursley gave himself a little shake and put **the cat** out of his mind.

Mandarin translation

It was on the corner of the street [...] **a cat** [...] Mr Dursley blinked and stared at **CAT**. **CAT** stared back.

As Mr Dursley drove around the corner and up the road, he watched **DEM. CAT** in his mirror. **CAT** was now reading the sign that said Privet Drive [...] Mr Dursley gave himself a little shake and put **CAT** out of his mind.

Going cross-linguistic II

It was on the corner of the street [...] **a cat** [...] Mr Dursley blinked and stared at **the cat**. **It** stared back.

As Mr Dursley drove around the corner and up the road, he watched **the cat** in his mirror. **It** was now reading the sign that said Privet Drive [...] Mr Dursley gave himself a little shake and put **the cat** out of his mind.

Hindi translation

It was on the corner of the street [...] **a cat** [...] Mr Dursley blinked and stared at **CAT**. **CAT** stared back.

As Mr Dursley drove around the corner and up the road, he watched **CAT** in his mirror. **CAT** was now reading the sign that said Privet Drive [...] Mr Dursley gave himself a little shake and put **CAT** out of his mind.

Going cross-linguistic II

It was on the corner of the street [...] **a cat** [...] Mr Dursley blinked and stared at **the cat**. **It** stared back.

As Mr Dursley drove around the corner and up the road, he watched **the cat** in his mirror. **It** was now reading the sign that said Privet Drive [...] Mr Dursley gave himself a little shake and put **the cat** out of his mind.

Russian translation

It was on the corner of the street [...] **a cat** [...] Mr Dursley blinked and stared at **CAT**. **CAT** stared back.

As Mr Dursley drove around the corner and up the road, he watched **CAT** in his mirror. **CAT** was now reading the sign that said Privet Drive [...] Mr Dursley gave himself a little shake and put **CAT** out of his mind.

Going cross-linguistic II

Interim conclusion

Our Mandarin data are fully in line with the intuition that demonstratives are preferred over bare nouns if the sentence in which the anaphor occurs is 'less tightly connected' to the sentence in which its antecedent occurs.

However, Mandarin turns out to occupy a special position in this respect and it is not possible to derive the Mandarin facts from the language-independent proposals of Simpson & Wu and Dayal & Jiang without making problematic predictions for other articleless languages like Hindi and Russian.

The contours of an analysis: step 1

- > Our cross-linguistic data suggest that the Mandarin demonstrative is doing more than a run-of-the-mill demonstrative should be doing and that – in so doing – limits the playing field of the bare noun in strong definiteness contexts (*Going cross-linguistic II*).
- > This leads me to assume that the Mandarin demonstrative has an extended use and does double duty as a demonstrative and a definite article.

The contours of an analysis: step 2

- > Assuming that the Mandarin demonstrative functions as a definite article in its extended use, we need to define the place it occupies in the typology of definite articles.
- > The traditional weak/strong typology is not rich enough. Our cross-linguistic data suggest that it's the strong side of the typology that has to differentiate between the uses that the demonstrative has taken over and the uses that remain with the bare noun (*Going cross-linguistic I*).

The contours of an analysis: step 3

- > Based on the fact that bare nouns in Mandarin, Hindi and Russian can all function as strong definites in 'tightly connected' stretches of discourse, I assume that this is the type of discourse for which we want to rely on standard dynamic machinery: indefinites introducing discourse referents and strong definites referring back to these.
- > This move entails that the standard dynamic machinery no longer applies to discourse stretches of arbitrary length but only to small substretches.

The contours of an analysis: step 3

$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$

The contours of an analysis: step 3

x_1, x_2

x_3

x_4, x_5, x_6, x_7

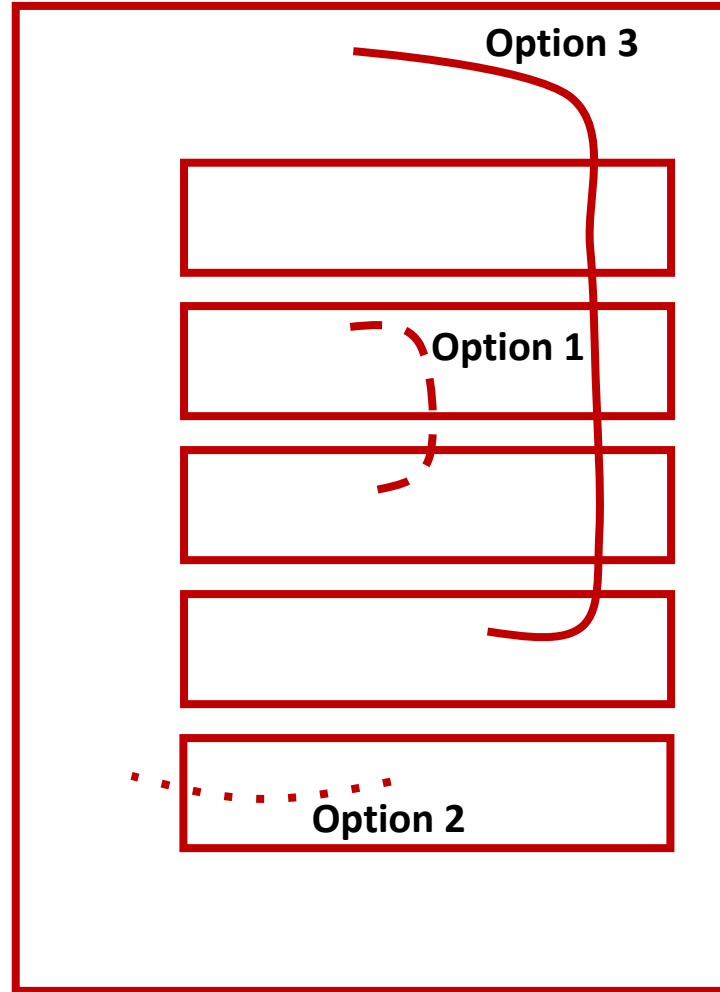
x_8, x_9

x_{10}

The contours of an analysis: step 4

- > The next question is how we should conceive of the role of the definite article uses of the demonstrative.
- > With the setup we have this far, there are (at least) three options:
 - (i) we can model them as granting access to referents in adjacent stretches of discourse or we can model them as granting access to
 - (ii) another parallel or a
 - (iii) hierarchically higher level.

The contours of an analysis: step 3



The contours of an analysis: step 4

- > Simpson & Wu's and Dayal & Jiang's intuition about the use of demonstratives in 'less tightly connected' stretches of discourse may seem to favor option 1.
- > However, Options 2 and 3 are more in line with current thinking about accessibility relations and are likely to be equally successful at capturing Simpson & Wu's and Dayal & Jiang's intuition.

The contours of an analysis: step 5

- > The final question is what kind of level can be hierarchically or parallel to the different stretches of discourse.
- > I'm interested in knowing how you feel about this level as involving a representation of the discourse-related common ground of speaker and addressee.
- > As for Mandarin demonstratives, my proposal is to analyze them as the grammaticalized way to access referents in this common ground, leading to the blocking of the bare noun. No such blocking would happen in Russian or Hindi, accounting for our data.

Definiteness in Mandarin: conclusion




- > Our Mandarin data are in line with the intuition that Mandarin demonstratives are preferred over bare nouns when they occur in a sentence that is 'less tightly connected' to the sentence introducing their antecedent.
- > This generalization differs from the one proposed by Jenks (2018) according to which bare nouns do not allow for anaphoric readings (see also *Going cross-linguistic I*) and the one proposed by Ahn (2019), according to which the presence of the demonstrative is optional.

Definiteness in Mandarin: conclusion

- > The fact that Russian and Hindi allow their bare nouns to occur in contexts that typically require the demonstrative in Mandarin shows that their distribution in Mandarin cannot be derived from more general considerations about demonstratives.
- > The analysis I propose is one in which the demonstrative has been recruited as a strong definite article and specializes in referring to referents in the common ground of speaker and addressee.

Talk

overview

- > Reports on the work of the **HHRM group**, with a focus on **Mandarin**.
- > Argues that:
 - For **indefiniteness**: Mandarin has a full-fledged indefinite article and that bare nouns with an indefinite interpretation are (pseudo-)incorporated; 
 - For **definiteness**: There are two types of strong definiteness, one relying on bare nouns, the other relying on the demonstrative; 
 - **Methodology-wise**: *Translation Mining* can insightfully be applied to study macro-variation. 

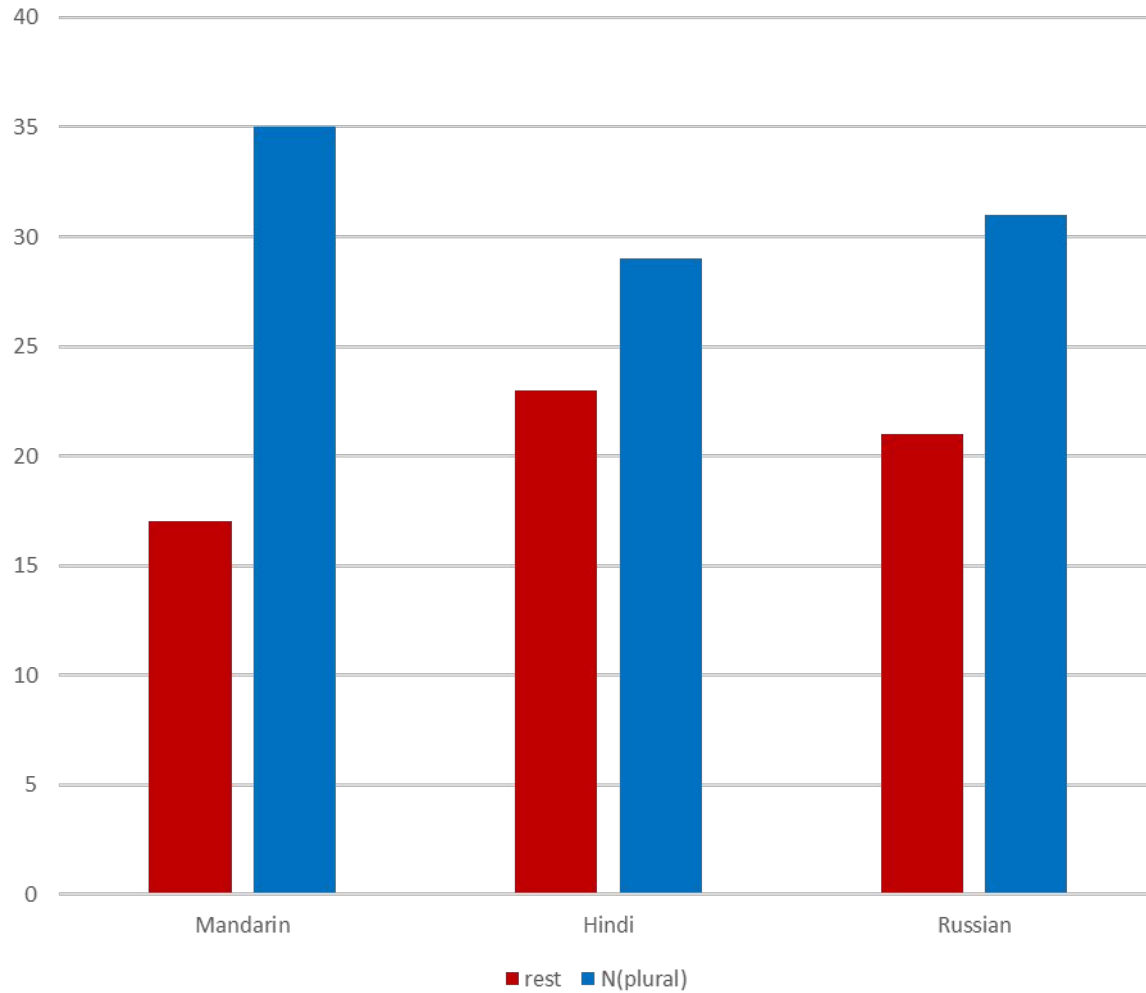
Thank you for your attention!

Study 1:

Results

Plural indefinite contexts

Distribution of BNs in Mandarin, Hindi and Russian



N_{plural}	rest	$N_{\text{(plural)}}$	total
Mandarin	17	35	52
Hindi	23	29	52
Russian	21	31	52

Fisher's Exact Test

Mandarin-Hindi: $p=0.3$

Mandarin-Russian: $p=0.5$

Hindi-Russian: $p=0.8$