Quaderni di Linguistica e Studi Orientali- QULSO
Working papers in Linguistics and Oriental Studies

Firenze University Press
https://oajournals.fupress.net/index.php/bsfm-qulso

Citation: G. Samo, F. Si (2022) Optionality of 的 *De* in Chinese Possessive Structures: A Quantitative Study. *Qulso* 8: pp. 37-53. doi: http://dx.doi.org/10.13128/QULSO-2421-7220-13602.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Optionality of 的 *De* in Chinese Possessive Structures: A Quantitative Study

Giuseppe Samo, Fuzhen Si
Beijing Language and Culture University
(samo@blcu.edu.cn; sifuzhen@blcu.edu.cn)

Abstract:

In this paper, we explore quantitative tools to investigate the optionality with respect to the presence of the possessive marker 的 *de* in alienable and inalienable constructions in Chinese. We explore three models to account for this optionality. The main model predicts syntactic nature to optionality, in which the syntactic structure plays a role in the licensing/or lack of licensing of the possessive marker, such as overtly realizing pragmatically defined contexts (e.g., the nature of the possessor) or discourse properties. We compare this model to two control groups. The first control group states that the behavior of the marker 的 *de* is highly dependent to the lexical properties of the possessum, while the second control group accounts for optionality as random, possibly given by chance. Corpus counts support that the syntactic model better captures the data. Finally, we discuss our results considering a cartographic approach.

Keywords: *Cartography, Chinese, Possessive Markers, Quantitative Syntax*

1. Introduction

A plethora of works in generative syntax has investigated the syntactic configurations of possessive structures and their interfaces at both lexical and semantic dimensions (Szabolcsi 1983; Cardinaletti 1998; Longobardi 2000; Haegeman 2004; Bernstein 2005; Si 2014, 2017 *inter alia*). Forms of asymmetries can be found with respect to the inalienable and alienable nature (Bickel and Nichols 2013) of the possessed element (*possessum*). Following Alexiadou, Haegeman and Stavrou (2007: 52), "alienable possession implies a possessor who is involved in acquiring the possessum, while inalienable possession is intrinsic, intimate possession which does not need to be acquired".

Let us discuss the differences by taking (1) and (2) as initial references, by comparing evidence from Chinese (1a, 2a) and English (1b, 2b).

(1) a. Chinese
我(的)妈妈
Wǒ de māma
I DE mother
'My mother'

b. English
My mother

(2) a. Chinese
我*/#(的)玩具
*Wǒ */#(de) wánjù*
I DE toy
'My toy'

b. English
My toy

Kinship, like 妈妈 *māma* 'mother' and *mother* in (1a, 1b), is tendentially interpreted as an inalienable possession, whereas other lexical elements such as 玩具 *wánjù* 'toy' and *toy* can be considered as encoding alienable possessions that can be acquired. The alienable and inalienable nature of the possessum seems marking dimensions of linguistic variability between English and Chinese. While English data (1b, 2b) do not superficially show asymmetries with respect to alienable and inalienable constructions, the Chinese possessive marker 的 *de* is described as optional in inalienable conditions (as in 1a), but obligatory in typical alienable constructions (2a, marked *) when uttered in isolation. However, 的 *de* in alienable constructions can be omitted, when additional conditions are met, syntactically and pragmatically (Si 2017; marked as # in 2a, more details will be discussed in the section 2). In the spirit of the program of syntacticisation of semantic and pragmatic properties (Cinque and Rizzi 2010; Si 2011; Rizzi and Cinque 2016), we will investigate whether the optionality is derived by the activation of functional projections within the syntactic architecture "relevant for the structuring of discourse" (Rizzi and Cinque 2016: 145).

In this paper, we aim to quantify this optionality and we solely focus on Chinese data. We explore three models to account for the optionality. The main model explores a syntactic nature to optionality, in which morpho-syntax (in terms of functional projection) plays a role in the licensing/or lack of licensing of the possessive marker, such as overtly realizing pragmatically defined contexts or discourse properties. We compare this model to two control groups, one lexical model, in which the behavior of the marker 的 *de* is highly dependent to the lexical properties of the possessum and a second control group stating the optionality is to be considered random, possibly given by chance. We will adopt quantitative methods in the spirit of the framework of Quantitative Computational Syntax (Merlo 2015, 2016; Merlo and Ouwayda 2018; Samo and Merlo 2019, 2021; Gulordava and Merlo 2020; Merlo and Samo 2022), using frequency as a dependent variable to test linguistic proposals. Specifically, we quantify the optionality in terms of frequency of lexical items with pronominal possessors in a large-scale database of heterogeneous sources of Chinese.

To reach this goal, we proceed as follows. In section 2, we introduce a syntactic proposal concerning the architecture of Chinese possessive structures within a cartographic framework.

In section 3 we quantify our models and our hypotheses. In section 4, we present materials and methods of our study. Section 5 shows results and discusses. Finally, section 6 concludes.

2. Some notes on the syntax of possessive structures in Chinese

The literature on the syntax of possessive structures is rich and has explored different populations of speakers and languages (Szabolcsi 1983, 1987, 1994; Cardinaletti 1998; Longobardi 2000; Gavrouseva 2000; Haegeman 2004; Matteini 2007; Alexiadou *et al.* 2007; Si 2014, 2017). Similarly, the relation between the possessor and the possessum has been deeply investigated (Bally 1926; Nichols 1988; Guéron 2006; Bickel and Nichols 2013; Rooryck 2022 *inter alia*).

Cross-linguistic variability emerges. For example, Alexiadou *et al.* (2007) consider English alienable and inalienable possessive structures "realized in syntactically identically ways, as pronominal genitives, possessive pronouns or post-nominal PPs" (*ibidem*, 552). On the other hand, Si (2014, 2017) postulates at least two heads are required to account for the asymmetries (such as those observed in 1 and 2) detectable in Chinese. An important factor for the stipulation of two heads is the optionality which is observable with inalienable possessive structures in kinship, part-whole relation, and locative possession (see also Ursini and Huang 2020: 5 for the latter point).

On the other hand, it is possible to find cases in which 的 *de* can be omitted in alienable constructions. For example, as given in (3), 的 *de* is optional when the possessive structure is embedded in a larger nominal phrase or relative clause (3c, d), but required when the possessive structure is uttered in isolation (3a, b).

(3) a. 那是什么？
Nà shì shénme
that is what
what is that?

b. * 我玩具
Wǒ wánjù
I toy
my toy

c. 我的玩具
Wǒ de wánjù
I de toy
my toy

d. 我玩具的颜色。
Wǒ wánjù de yánsè
I toy DE color
'The color of my toy'

e. 我的玩具的颜色
Wǒ de wánjù de yánsè
I de toy DE color
'The color of my toy'

When a possessive structure used as an embedded phrase of a larger structure (3c, d), the possessive marker 的 *de* within the embedded possessive becomes optional and both types of possessives behave similarly. In brief, the use of 的 *de* in alienable possessive is required when they are independently used as a fragment structure in a context; omission of 的 *de* in other situations is restricted.

The main arguments supporting at least two positions for Chinese are listed in (4) extracted from (Si 2017: 198ff).

(4) Adapted from (Si 2017: 198)
1. More than one position is available for the licensing of possessors;
2. Possessors are base-merged inside the lexical layer and may target different positions within the DP;
3. Alienable possessive relation might be marked higher than inalienable possession in terms of their base position;
4. *De* (的) can be a PossP marker (functional head) in both types of possessive structures.
5. *De* (的) in both type of possessive structures is pragmatically/discourse relevant or contextually determined.

Further empirical evidence licenses the postulation of at least two different positions. For example, copular constructions are only allowed in alienable-like possessions. Similarly, only alienable-like possessive elements can target object positions in double object constructions. As additional evidence, Si (2017: 204-206) explores a series of variety across China. Of a particular interest is the variety of Jiaocheng dialect (a Jin Dialect, spoken in Shanxi province, Northern China), which shows morphological change on the personal pronominal possessor according to the inalienable-like/alienable-like status of the possessum. The pronominal status will be also investigated in this paper from a quantitative point of view.

Based on Si (2017: 210), a cartography of possessives in Chinese is presented in (5).

(5) $[_{\text{SpecPossP}}$ Possessor $[_{\text{Poss'}}$ (的) $[_{\text{SpecModificationP}}[_{\text{Modification}}$ <Possessor> $[_{\text{NP}}$ Possessum]]]
{Contextually determined possession {Lexically determined modificational relation}}

As a matter of fact, the goal of cartographic studies is to provide a powerful analytical tool being able to capture microvariability intra-linguistically and cross-linguistically based on syntactic architecture descriptions resulting from the interaction of fine-grained maps of syntactic configurations and basic computational operations (Rizzi 2004, 2015a, 2015b; Samo 2019a: ch. 1).

The heuristic capacity of cartographic studies offer transparent configurations to which interpretive routines can be applied resulting in the syntacticisation of semantic, pragmatic, and prosodic properties (Cinque and Rizzi 2010; Si 2011; Rizzi and Cinque 2016).

Syntactically speaking, the constructions without the possessive marker 的 *de* would be an unmarked or default construction, marking intrinsic relation, while constructions with 的 *de* would be the representations of a peripheral-pragmatic position, and the 的 *de* is a functional head leading to syntacticisation of the contextually determined possessive relation.

In a nutshell, we aim to detect whether the optionality is triggered by syntax. We compare this model against two models stating that optionality may results from a pure lexical factor and from a model assigning random factors. As it will be discussed in section 4, we operate our quantitative study relying on pronominal possessors which facilitate and enable fully automatized retrieval in large-scale (partially, non-annotated) corpora.

Our measures will be (raw) frequency and distributions. The interaction between frequency of grammatical structures and theory of grammar has been highly debated in the literature (Bresnan *et al.* 2001; Yang 2004; Ibbotson 2013; Yang *et al.* 2017 *inter alia*). In this work, we follow the guidelines of Quantitative Computational Syntax (Merlo 2016 and related works). Large-scale datasets (Nivre 2015 *inter alia*) allow "us to develop investigations of the correlation between quantitative linguistic properties and theory-driven abstract linguistic representations and operations" (Samo and Merlo 2021: 29). In other words, in the spirit of Merlo (2015, 2016), we adopt quantitative measures, such as frequency, as a dependent variable to test linguistic proposals.

3. Modelling theories on grammatical clauses

We explore a large-scale dataset of Mandarin Chinese, namely the BCC corpus (BLCU CHINESE CORPUS, 15 billion characters; created by Endong Xun, Beijing Language and Culture University), by extracting the occurrences of nominal elements following pronominal possessors (all the queries will be presented in section 4.1.).[1]

Our measures will be represented by the raw frequency of lexical elements appearing in both forms and will represent the values of a bi-dimensional vectorial representation. The first dimension is the set of lexical items in possessive structures directly following the pronominal possessor, thus without the possessive marker 的 *de* (henceforth, Empty constructions, EMPTY) such as the phrase 我父亲 *wǒ fùqīn* 'my father'. The second dimension is given by the raw frequencies in which the same lexical element acting as a possessum is preceded by the possessive marker 的 *de* (henceforth, *de*-constructions, DE), e.g. 我的父亲 *wǒ de fùqīn* 'my father'.

In a nutshell, our datapoints would be a set of bidimensional vectors whose values are given by (i) the raw frequency of the possessum directly preceded by the pronominal form (empty) and (ii) the raw frequency of the possessum preceded by the possessive marker 的 *de*. For example, if the word e.g.: 父亲 *fùqīn* 'father' naturally occurs in the corpus investigation in 150 structures following a personal pronoun directly without 的 *de* (EMPTY) and 100 occurrences with the marker 的 *de,* the lexical entry 父亲 *fùqīn* 'father' will be encoded as the bi-dimensional vector 父亲 = (150, 100).

This work focuses on the testing, from a frequency point of view, of the program of syntacticisation (Cinque and Rizzi 2010; Si 2011; Rizzi and Cinque 2016). Cartographically marked order in grammatical clauses extracted from corpora (Samo 2019b, Samo and Merlo 2019, 2021) appears to be less frequent than expected than canonical sentences.

The inalienable/alienable nature can be thought as a pragmatically defined movement. Frequent orders will represent an "unmarked" option, while the less frequent one would be the representation of a peripheral position (such as 的 *de*, following Si 2017) in the vP, leading to syntacticisation. How can we account for a "preference"? We adopt the model discussed in Samo (2021) which considered "preference" of a pattern over another as the distribution compared to a binomial probability. Following the syntactic derivation in (5), optionality is created by the syntactic configuration and the movement from a base-generation position to a higher position in the structure.

[1] *BCC* <http://bcc.blcu.edu.cn/>(07/2022). See also Xu (2015: 219, 243).

We explore several theories to account for the optionality. A first model involves a syntactic explanation and represents our main hypothesis. The definition of this model is given in M_{syn} and its hypothesis is stated in H_{syn}.

M_{syn}: The optionality is caused by the syntactic nature of the tree and by movement from a "canonical" external merge position to a pragmatically marked internal merge position.
H_{syn}: If optionality is syntactically given, we should observe a general "preference" in a direction, possibly towards the unmarked word order.

We compare the theory discussed in M_{syn} with two control groups. The first control group is related to the lexical nature of the possessum. If the optionality is given by the lexical form of the possessum and its semantics, then, for every lexical entry, we should observe that there should be a clear preference for one of the dimensions, either EMPTY or DE. According to this model, no role is played by syntax in explaining the optionality, but the presence/absence of the possessive marker 的 *de* is selected at the lexical level. This model is defined in M_{lex} and hypothesis H_{lex}.

M_{lex}: The optionality is caused by the nature of the lexical element.
H_{lex}: One of the dimensions for every collected datapoint x = (EMPTY, DE) should be equal to 0.

A more refined model, however, would imply that the alienable vs. inalienable status would play a role. In other words, an inalienable possessum should have a strong preference for presence of the possessive marker 的 *de* and no preference for EMPTY, while an inalienable possessive should show a chance-level optionality. However, at this stage, we cannot automatically retrieve the relevant properties, since the task of assigning a value as alienable and inalienable a priori does not represent an easy task for a set of lexical entries (see the detailed discussion in section 5). However, this model would be the opposite pattern of H_{syn} since it would predict an overall general preference for 的 *de* (both inalienable and alienable can have the possessive marker, but only inalienable can also be present in EMPTY constructions).

Finally, as a second control group we build a third model related to chance level. This simple, but non-trivial, model relies on the fact that the optionality might not be given by a syntactic configuration, but the presence/absence of 的 *de* in the construction is considered as a purely random factor. This theory, and the relevant hypothesis H_{rand} will act as a null hypothesis for the following two theories and hypotheses. The model is given in M_{rand} and its hypothesis in H_{rand}. This model will also serve as a control group.

M_{rand}: The optionality is random.
H_{rand}: The distributions of the two orders/dimensions for every collected datapoint x = (EMPTY, DE) is at a chance level.

We will present materials and methods in section 4.

4. Materials & Methods

We explored the Chinese corpus of the BCC (BLCU CHINESE CORPUS, 15 billion characters) which contains texts in Chinese from a heterogeneity of sources, mainly news and literature. Our sample of sentences, to have a fully automatized retrieval process, will only contain the possessor in a pronominal form. We therefore created selected queries in which the pronominal element/the pronominal element and the possessive marker are followed by *n,* which is the

standard annotation for nominal entities.[2] Table 1 presents queries, number of tokens retrieved and a naturally occurring example for each query.

Query	Gloss	Tokens	Naturally occurring example
我n	'I + Noun'	1,005,060	我爸说, […] *wǒ bà shuō* 'my father said'
我的n	'I + DE + Noun'	609,509	我的朋友也这样想 *Wǒ de péngyǒu yě zhèyàng xiǎng* 'my friends think so too'
你n	'you.SING + Noun'	505,660	你大哥正在运功 *Nǐ dàgē zhèngzài yùn gōng* 'Your big brother is practicing Qi Gong'
你的n	'you.SING + DE + Noun'	359,341	我从你的脸孔中认出了她 *Wǒ cóng nǐ de liǎn kǒng zhōng rèn chūle tā* 'I recognized her from your face'
他n	'he + Noun'	608,629	他父亲叫他"小崽子" *Tā fùqīn jiào tā "xiǎo zǎizi"* 'His father called him "little boy"'
他的n	'he + DE + Noun'	566,763	他的世兄一定走了 *Tā de shìxiōng yídìng zǒule* 'His brother must have been gone'
她n	'she + Noun'	246,943	他拉住她手道 *Tā lā zhù tā shǒu dào* 'He took her hand and said'

[2] This query, naturally, can detect cases in which the noun is preceded by a pronominal form which is inserted in a dative (double object construction, see Si 2021 for Chinese) of the type of the naturally occurring example 你会帮我们大忙，对不？*Nǐ huì bāng wǒmen dàmáng, duì bù?* 'You'll do us a big favor, won't you?'. Another case could be represented when two linearly adjacent elements have no any sectional relation from syntactic point of view, although semantically speaking they might have a possessive relation. Let us observe the naturally occurring example extracted from the corpus under investigation, given in (i).

(i) 他面色阴郁地盯着床单
Tā miànsè yīnyù de dīng-zhe chuángdān
He face-color gloomy DE stare-ing at sheets
'He was staring gloomily at the sheets'

In (i), 他 *tā* 'he' and 面色 *miànsè* 'face-color' are linearly adjacent. While semantically the two words hold a possessive relation, they do not syntactically really form up any structure, for that 面色阴郁 *miànsè yīnyù* 'face-color gloomy' is an adverbial of the predicate 盯着床 *dīngzhe chuángdān* 'stare at sheets'. We were not able to manually operate on the entire dataset, but these examples represented, from a sample that we manually investigated, an extremely reduced portion that we consider unable to bias our results from "unwanted" tokens.

她的n	'she + DE + Noun'	360,836	她的笑容为什么可以这么灿烂、温暖 ? *Tā de xiàoróng wèishénme kěyi zhème cànlàn, wēnnuǎn ?* 'Why is her smile so bright and warm?'
它n	'it + Noun'	87,564	它妈妈是白色长毛的纯波斯种 *Tā māmā shì báisè chángmáo de chún bōsī zhǒng* 'Its mother is a pure Persian with long white hair'
它的n	'it + DE + Noun'	77,785	它的屏幕实在出色 *Tā de píngmù shízài chūsè* 'its screen is really good'
我们n	'we + Noun'	207,211	今天一天都在我们家门口走来走去 *Jīntiān yìtiān dōu zài wǒmen jiā ménkǒu zǒu lái zǒu qù* 'Walked around our door all day today'
我们的n	'we + DE + Noun'	152,577	那又不是我们的错 ! *Nà yòu búshì wǒmen de cuò* ! 'It's not our fault!'
你们n	'you.PLUR + Noun'	42,117	你们老板他老这个样子吗? *Nǐmen lǎobǎn tā lǎo zhège yàngzi ma ?* 'Is your boss always like this?'
你们的n	'you.PLUR + DE + Noun'	27,912	你们的男朋友合格了吗 *Nǐmen de nánpéngyǒu hégé le ma ?* 'Are your boyfriends qualified?'
他们n	'they.MASC + Noun'	64,948	因为他们爹妈是80后 ! *Yīnwèi tāmen diēmā shì 80 hòu* ! 'Because their parents are born in the 80s!'
他们的n	'they.MASC + DE + Noun'	121,629	而他们的收入何尝不比社会平均收入水平高出许多。 *Ér tāmen de shōurù hécháng bùbǐ shèhuì píngjūn shōurù shuǐpíng gāo chū xǔduō* 'And their income is much higher than the social average income level'
她们n	'they.FEM + Noun'	7,946	她们公司超好 *Tāmen gōngsī chāo hǎo,* Their company is super good,
她们的n	'they.FEM + DE + Noun'	10,984	她们的歌声还是如此美妙。 *Tāmen de gēshēng háishì rúcǐ měimiào.* Their voices are still so beautiful.

它们n	'they.NON-HUM + Noun'	3,794	它们祖上是远亲 *Tāmen zǔshàng shì yuǎnqīn* 'Their ancestors are distant relatives
它们的n	'they.NON-HUM + DE + Noun'	13,343	欣赏什么呢？它们的风味。 *Xīnshǎng shénme ne? Tāmen de fēngwèi.* 'What do you appreciate? their flavor.'

Table 1. Query, gloss, number of tokens and a naturally occurring example.

The query also detected cases of embedded possessive structures such as the naturally occurring example 我们校训的第一条就是"酷爱读书" *Wǒmen xiàoxùn de dì yī tiáo jiùshì "kù'ài dúshū"* "The first rule of our school motto is "Love reading". These complex structures represent tokens, since they can be compared to sentences of the type (containing the optional 的 *de*) 我们的校训的第一条就是 "酷爱读书 *Wǒmen de xiàoxùn de dì yī tiáo jiùshì "kù'ài dúshū"*.

The interface at the BCC corpus only allows to export 10,000 results in a relevant *.txt* format. Based on the first 10,000 we operated our counts by running a frequency function using R (R development team, 2022). Our counts, as mentioned in section 3, represent the coordinates of our datapoints.[3] The first dimension is the raw frequencies in which every type of lexical element co-occurs in structures directly preceded by the pronominal form, while the second is the raw frequencies of possessions inserted in a DE construction.

The actual distribution and their size will be compared with two fictional distributions given by the postulated control group by M_{lex} and M_{rand}. The group of distributions representing M_{lex} will assign to every lexical entry a distribution of 1 or 0 according to the highest frequency of one of the dimensions. The group of distributions representing M_{rand} would assign a 0.50 distribution to both dimensions. As an example, let us take the actual datapoints (see section 5) with respect to 父亲 *fùqīn* 'father' = (350, 272). Transforming the values terms of distributions, we have 父亲 *fùqīn* 'father' = (0.562, 0.437). The distribution for the control groups M_{lex} and M_{rand} would be respectively 父亲 *fùqīn* 'father' = (1.000,0.000) and 父亲 *fùqīn* 'father' = (0.500,0.500).

Results and discussions are presented in section 5.

5. Results & Discussions

Out of 14139 lexical entries/datapoints, only a subset of nominal elements (7954 types, 56%) show at least one occurrence pro-each dimension. Figure 1 shows the data point distribution.

[3] All data are available at the following link <https://github.com/samo-g/deoptionality.git> (07/2022).

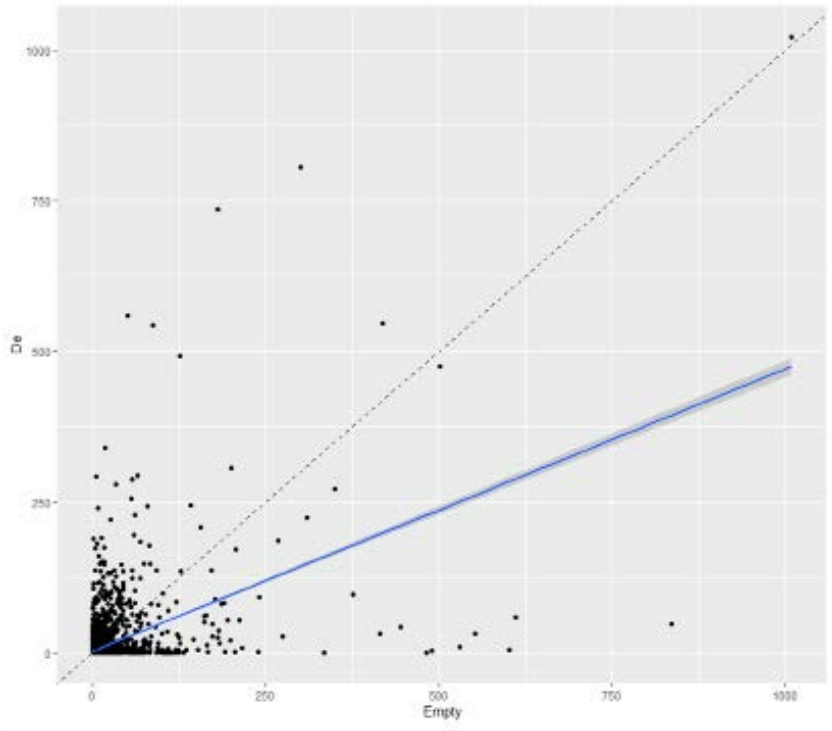

Figure 1. Distribution of datapoints (*n = 14,139*). Raw Frequency of constructions without (EMPTY, x axis) and with possessive marker 的 *de* (DE, y axis). The two axes would be a reference for M_{lex}, the dashed line represents M_{rand} and the blue line is the actual linear regression.

First of all, our datapoints are statistically significant from a distribution predicted by the lexical model M_{lex} ($t(28276) = 13.4394$, $p < 0.0001$) and from the distribution predicted by the random model M_{rand} ($t(28276) = 2.2515$, $p < 0.05$). As expected, we observe a tendency towards the unmarked configuration ($R^2 = 0.24017$). This difference can be due that the optionality is therefore created by the syntactic context in which the possessive structure occurs.

The syntactic direction can be accounted for in terms of preference. We define preference as the dimension with the locally higher raw frequency. To detect whether the preference is given by chance, we run a binomial test (see also Samo 2021 and reference therein). The binomial test gives us the probability of *k* successes (the number of occurrences of the highest dimension) in *N* independent trials (the number of occurrences of the lexical item), given a base probability *p* (0.50). We test it on the 100 most frequent lexical entries which have at least one occurrence pro-dimension. We then divided these 100 most frequent lexical entries according to their potential status as alienable or inalienable possessives. All the relevant data are given respectively in Table 3 (54 lexical entries) and Table 4 (32 lexical entries) in the Appendix.[4]

[4] We manually investigated the results. Out of the 100 most frequent words we removed fourteen possible noises given by the annotation. We removed words that are clearly case of appositive configurations, e.g. 人 *rén* 'people' (rank 3), 母女 *mǔ nǚ* 'mother and daughter' (rank 17), 个人 *gèrén* 'personal' (rank 60), 夫妻 *fūqī* 'couple' (rank 86), possible verb forms annotated as nouns, e.g. 代表 *dàibiǎo* 'represent (rank 21), 感觉 *gǎnjué* 'feel' (rank 36), 建

Our results can be read as follows. In Table 3 (Appendix), we observe that 的 *de* is preferred options in alienable constructions (78% types (42 out of 54); 65% in terms of tokens (9525 out of 14628, *binomial p < .000001*). DE is a disfavoured option inalienable possessives (Table 4 in Appendix) in terms of types (11 out of 32) and tokens (5288 out of 13099, *binomial p < .000001*). The results in the Table 4 (Appendix) perfectly proves the assumption that the occurrence of 的 *de* also in inalienable possessives is "optional" according to the pragmatically defined context as what can be observed by the higher frequencies of constructions with 的 *de* (e.g. 眼睛 *yǎnjīng* 'eye', 孩子 *háizi* 'child').

The results in Table 3 (Appendix) needs more explanation, although the preference of using 的 *de* in the alienable possessives seems obvious. What we do need to explain under which condition 的 *de* can also be omitted in alienable possessive structures and if this omission can be comparable to the cases in which 的 *de* can be omitted in inalienable possessives. In a nutshell, we need to understand whether the omission of 的 *de* in the inalienable possessives does constitute a challenge for our hypothesis on the syntactic distinction between the two types of possessives.

A case of existing omission of 的 *de* is when the possessive relation is based on an affiliation, as in 我们公司 *wǒmen gōngsī* 'our company', 我们国家 *wǒmen guójiā* 'our country', 我们单位 *wǒmen dānwèi* 'our unit'. In this case, plural possessors are preferred. This is what we quantitively observe. In Table 2, we compare singular and plural pronominal forms (我 *wǒ* 'I', 你 *nǐ* 'you.sing', 他/她 *tā* 'he/she' 我们 *wǒmen* 'we', 们 *nǐmen* 'you.plur', 他们/她们 *tāmen* 'they') with respect to the frequency of 公司 *gōngsī* 'company' and 国家 *guójiā* 'country'. As Table 2 shows, the raw frequency in plural form is higher than in singular forms (公司 *gōngsī*: 409 out of 426, *binomial p < .000001*; 国家 *guójiā* 'country': 286 out of 287, *binomial p* < .000001).[5]

It has also been noticed that in quite a few of Chinese dialects the plural markers can serve as a marker of possession (see detailed discussion in Si 2017).

	1st sing.	2nd sing.	3rd sing.	1st plur.	2nd plur.	3rd plur.
公司 *gōngsī* 'company'	9	5	3	127	225	57
国家 *guójiā* 'country'	1	0	0	218	53	15

Table 2. Raw frequencies of 公司 *gōngsī* 'company' and 国家 *guójiā* 'country' in the different persons in EMPTY constructions.

议 *jiànyì* 'suggestion' (rank 43), 计划 *jìhuà* 'plan' (rank 57) and 信 xìn 'letter/trust' (rank 51), as well as elements in which it is quite impossible to provide a clear semantics of alienabilty/inalienability such as 丈夫 *zhàngfū* 'husband' (rank 22), 妻子 *qīzi* 'wife' (rank 68), 一生 *yīshēng* 'lifetime' (rank 35), 省 *shěng* 'province' (rank 75) or unclear such as 时候 *shíhòu* 'when' (rank 44).

[5] Additional evidence for what we state comes from the comparison with respect to number (singular, plural) in 的 *de* constructions' data of 公司 *gōngsī* 'company' and 国家 *guójiā* 'country'. The distributions for 国家 *guójiā* with plural possessors are similar (92% of plural data), but smaller to the discussed configurations in Table 2 (99%) in empty configurations. The set of distributions in number of the possessor with 公司 *gōngsī* 'company' are totally different, and in line with what it is proposed here. In 的 *de* constructions, singular possessors represent the 56,1% of tokens and plural possessors the 43,9%, while in empty constructions there is a strong preference for plural possessors (96%), similarly to 国家 *guójiā* 'country'.

The quantitative analysis discussed here has demonstrated that there is no clear mapping between the usage of 的 *de* and the nature of the lexical entry. From a theoretically point of view, the functional projection hosting 的 *de* might be described as a locus of first internally merge, after elements have been retrieved from the lexicon (see Rizzi 2016 for discussions in terms of cartography), while the omission could be derived, as discussed in Si (2017) via movement and by a series of instructions triggered by the relevant functional projections (plausibly, e.g., in the spirit of Rizzi 2017).

6. Conclusions

In this work we aimed to quantify the optionality, by exploring an automatic retrieval of possessive structures from a large-scale database. We built three models: a syntactic, a purely lexical and a fully random model. The distributions of our datapoints, intended as the raw frequency of nominal entities in pronominal possessive structures in which we observe the presence/absence of the possessive marker 的 *de*. What we have found is that a syntactic model better explains the data.

Empirically speaking, we still observe some forms of lexical choices. The asymmetries with respect to 的 *de* in the two types of possessive structures is clear. Methodologically speaking, a combination of the quantitative models and theory should be pursued in uncovering the fine-grained syntactic mechanisms. Quantitative methods in cartography would provide further insights with respect to grammatical structures, but, we believe, they must be always together with theoretical guidance and support. Future studies should refine the methodology, such as investigating syntactically annotated corpora, to observe, if any, differences between different types of possessors or their syntactic position within the syntactic architecture (e.g., subject, object).

References

Alexiadou, Artemis, Liliane Haegeman and Melita Stavrou. 2007. *Noun Phrase in the Generative Perspective*. Berlin: Mouton de Gruyter.

Bally, Charles.1926. "L'expression des idées de sphère personnelle et de solidarité dans les langues indo-européennes". In *Festschrift Louis Gauchat*, hrsgg. von Franz Frankhauser und Jakob Jud, 68-78. Aarau: Sauerländer.

Bernstein, Judy B. 2005. "On the morpho-syntax of possessive constructions". *Recherches linguistiques de Vincennes* 34: 55-76.

Bickel, Balthasar & Johanna Nichols. 2013. "Obligatory Possessive Inflection". In *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute For Evolutionary Anthropology. <https://wals.info/chapter/58> (07/2022).

Bresnan, Joan, Shipra Dingare and Christopher D. Manning. 2001. "Soft constraints mirror hard constraints: Voice and person in English and Lummi". In *Proceedings of the LFG0 conference*, ed. by Miriam Butt and Tracy H. King, 13-32. Stanford: CSLI Publications.

Cardinaletti, Anna. 1998. "On the deficient/strong Opposition in Possessive Systems", in *Possessors, Predicates and Movement in the Determiner Phrase*, ed. by Artemis Alexiadou and Chris Wilder, 17-54. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Cinque, Guglielmo and Luigi Rizzi. 2010. "The cartography of syntactic structures". In *Oxford Handbook of Linguistic Analysis*, ed. by Bernd Heine and Heiko Narrog, 51-65. New York: Oxford UP.

Gavruseva, Elena. 2000. "On the syntax of possessor extraction", *Lingua* 110: 830-772.

Guéron, Jacqueline. 2006. "Inalienable possession". In *The Blackwell Companion to Syntax*, ed. by Martin Everaert and Henk van Riemsdijk, 589-638. Malden, MA: Blackwell Publishing.

Gulordava, Kristina and Paola Merlo (2020). "Computational quantitative syntax: The case of Universal 18." In *Romance Languages and Linguistic Theory 16: Selected papers from the 47th Linguistic Symposium on Romance Languages (LSRL), Newark, Delaware*, ed. by Irene Vogel, 109-132. Amsterdam: John Benjamins.

Haegeman, Liliane. 2004. "DP-Periphery and Clausal Periphery: Possessor Doubling in West Flemish". In *Peripheries, Syntactic Edges and their Effects,* ed. by David Adger, Catherine De Cat and George Tsoulas, 211-240. Dordrecht-Boston-London: Kluwer Academic Publishers.

Ibbotson, Paul. 2013. "The Scope of Usage-Based Theory". *Frontiers in Psychology* 4 (255): 1-15.

Longobardi, Giuseppe. 2000. "Postverbal Subjects and the Mapping Hypothesis". *Linguistic Inquiry* 31 (4): 691-702.

Matteini, Simona. 2007. O*n the morphosyntactic acquisition of the German DP by Italian adult L2 Learners: Possessive Constructions, Noun Placement, and Inflectional Morphology on Nominal Modifiers*. PhD Thesis. Florence: University of Florence.

Merlo, Paola. 2015. "Predicting word order universals". *Journal of Language Modelling* 3(2): 317-344.

Merlo, Paola. 2016. "Quantitative Computational Syntax: some initial results". *Italian Journal of Computational Linguistics* 2(1): 11-30.

Merlo, Paola and Sarah Ouwayda. 2018. "Movement and structure effects on universal 20 word order frequencies: a quantitative study". *Glossa* 3 (84): 1-35.

Merlo, Paola and Giuseppe Samo. 2022. "Exploring T3 Languages with quantitative computational syntax", *Theoretical Linguistics* 48: 1-11.

Nichols, Johanna. 1988. "On alienable and inalienable possession". In *In honor of Mary Haas: From the Haas Festival Conference on Native American linguistics*, ed. by William Shipley, 557-609. New York: Mouton de Gruyter.

Nivre, Joakim. 2015. "Towards a Universal Grammar for Natural Language Processing". In *International Conference on Intelligent Text Processing and Computational Linguistics: 16th International Conference*, *CICLing 2015, Proceedings, Part I*, (Cairo, Egypt, April 14-20, 2015), ed. by Alexander Gelbukh, 3-16. Cham: Springer.

R Core Team. 2021. *The R Project for Statistical Computing*. <https://www.R-project.org/> (07/2022).

Rizzi, Luigi. 2004 "Locality and Left Periphery", In *Structures and Beyond. The Cartography of Syntactic Structures*, vol. 3, ed. by Adriana Belletti, 223-251. Oxford: Oxford University Press.

Rizzi, Luigi. 2015a. "Cartography, Criteria, and Labeling". In *Beyond Functional Sequence. The Cartography of Syntactic Structures*, vol. 10, ed. by Ur Shlonsky, 314-338. Oxford-New York: Oxford University Press.

Rizzi, Luigi. 2015b. "Notes on Labeling and Subject Positions". In *Structures, Strategies and beyond – Studies in Honour of Adriana Belletti*, ed. by Elisa di Domenico, Cornelia Hamann and Simona Matteini, 17-46. Amsterdam: John Benjamins.

Rizzi, Luigi. 2016. "Monkey Morpho-Syntax and Merge-Based Systems." *Theoretical Linguistics* 42 (1-2): 139-145.

Rizzi, Luigi. 2017. "On the format and locus of parameters: The role of morphosyntactic features". *Linguistic Analysis* 41: 159–191.

Rizzi, Luigi and Guglielmo Cinque. 2016. "Functional Categories and Syntactic Theory", *Annual Review of Linguistics* 2: 139-163.

Rooryck, Johan. 2022. Reconsidering inalienable possession with definite determiners in French. *Isogloss. Open Journal of Romance Linguistics 8* (1)/5: 1-34.

Samo, Giuseppe. 2019a. *A criterial approach to the cartography of V2*. Amsterdam: John Benjamins.

Samo, Giuseppe. 2019b. "Cartography and Locality in German: a Quantitative Study with Dependency Structures." *Rivista Di Grammatica Generativa/Research in Generative Grammar* 41 (5): 1-26.

Samo, Giuseppe. 2021. "N-merge systems in adult and child grammars: a quantitative study on external arguments." *Qulso: Quaderni di Linguistica e Studi Orientali* 7: 103-130.

Samo, Giuseppe and Paola Merlo. 2019. "Intervention effects in object relatives in English and Italian: a study in quantitative computational syntax." In *Proceedings of the First Workshop on Quantitative Syntax (Quasy, SyntaxFest 2019)*, ed. by Xinying Chen and Ramon Ferrer-i-Cancho, 46-56. Paris: Association for Computational Linguistics.

Samo, Giuseppe and Paola Merlo. 2021. "Intervention effects in clefts: a study in quantitative computational syntax", *Glossa: a journal of general linguistics* 6 (1): 1-32.

Si, Fuzhen. 2011. "Syntacticization of pragmatic information at sentential and noun phrase levels (语用信息在句子和名词短语层面的句法化)", *Journal of Yili Normal University (Social Science Edition)* [伊犁师范学院学报（社会科学版）] (2): 92-5.

Si, Fuzhen. 2014. "*Zhao Benshan de Yeye and Zhao Benshan de Maozi*: Two Types of Possessive Structures in Chinese (赵本山的爷爷 和 赵本山的帽子 – 漫谈汉语中的两种领属结构)", *Journal of Language and Teaching Research* (语言教学与研究) 2 (2): 43-51.

Si, Fuzhen. 2017. "Two Types of Chinese Possessive Structures". In *Studies on Syntactic Cartography,* ed. by Fuzhen Si, 197-217. Beijing: China Social Sciences Press.

Si, Fuzhen. 2021. "Towards a cartography of light verbs". In *Current Issues in Syntactic Cartography: A crosslinguistic perspective*, ed. by Fuzhen Si and Luigi Rizzi, 217-242. Amsterdam: John Benjamins.

Szabolcsi, Anna. 1983. "The Possessor that Ran Away from Home". *The Linguistic Review* 3: 89-102.

Szabolcsi, Anna.1987. "Functional Categories in the Noun Phrase". In *Approaches to Hungarian*, vol. 2, ed. by István Kenesei, 167-190. Szeged: Jate.

Szabolcsi, Anna. 1994. "The Noun Phrase". *The Syntactic Structure of Hungarian Syntax and Semantics*, vol. 27, ed. by Ferenc Kiefer, and Katalin É. Kiss, 179-274. New York (NY): Academic Press.

Ursini, Francesco-Alessio and Aijun Huang. 2020. "The anaphoric properties of spatial nouns and prepositions in Mandarin: A formal account". *Lingua* 248: 1-27.

Xu, Jiajin. 2015. "Corpus-based Chinese studies: A historical review from the 1920s to the present". *Chinese Language and Discourse* 6 (2): 218-244.

Yang, Charles D. 2004. "Universal Grammar, Statistics or Both?". *TRENDS in Cognitive Sciences* 8 (10): 451-456.

Yang, Charles D., Stephen Crain, Robert Berwick, Noam Chomsky and Johan Bolhuis. 2017. "The growth of language: Universal Grammar, experience, and principles of computation". *Neuroscience & Biobehavioral Reviews* 81: 103-119.

Appendix

Rank	*Noun*	*Pinyin*	Gloss	Empty	DE	*n*	Preference	*binomial P*
2	心	*xīn*	heart	301	805	1106	DE	0.000000000
5	话	*huà*	talk	181	736	917	DE	0.000000000
8	声音	*shēngyīn*	sound	87	544	631	DE	0.000000000
11	名字	*míngzì*	name	51	560	611	DE	0.000000000
13	学校	*xuéxiào*	school	552	32	584	Empty	0.000000000
14	班	*bān*	class	530	9	539	Empty	0.000000000
18	公司	*gōngsī*	company	446	42	488	Empty	0.000000000
19	国家	*guójiā*	nation	377	97	474	Empty	0.000000000
20	关系	*guānxì*	relation	269	185	454	Empty	0.000015358

25	朋友	*péngyǒu*	friend	65	294	359	DE	0.000000000
26	意思	*yìsi*	mean	19	340	359	DE	0.000000000
27	事	*shì*	thing	57	288	345	DE	0.000000000
29	主人	*zhǔrén*	owner	80	243	323	DE	0.000000000
33	特点	*tèdiǎn*	features	5	293	298	DE	0.000000000
34	作用	*zuòyòng*	effect	62	228	290	DE	0.000000000
41	目光	*mùguāng*	look	60	196	256	DE	0.000000000
42	思想	*sīxiǎng*	thought	69	184	253	DE	0.000000000
45	世界	*shìjiè*	world	26	220	246	DE	0.000000000
47	心情	*xīnqíng*	feeling	84	148	232	DE	0.000007111
53	感情	*gǎnqíng*	emotion	75	147	222	DE	0.000000388
54	宿舍	*sùshè*	dormitory	200	21	221	Empty	0.000000000
58	目的	*mùdì*	purpose	13	191	204	DE	0.000000000
62	人生	*rénshēng*	life	58	137	195	DE	0.000000004
63	精神	*jīngshén*	spirit	69	123	192	DE	0.000027209
65	天	*tiān*	sky	2	189	191	DE	0.000000000
66	表情	*biǎoqíng*	expression	42	146	188	DE	0.000000000
67	目标	*mùbiāo*	target	7	181	188	DE	0.000000000
69	产品	*chǎnpǐn*	product	45	135	180	DE	0.000000000
72	优点	*yōudiǎn*	advantage	4	173	177	DE	0.000000000
73	脸色	*liǎnsè*	complex-ion	95	79	174	Empty	0.029036583
74	行为	*xíngwéi*	behavior	10	161	171	DE	0.000000000
75	省	*shěng*	province	167	1	168	Empty	0.000000000
77	要求	*yāoqiú*	require	31	136	167	DE	0.000000000
78	爱情	àiqíng	love	17	150	167	DE	0.000000000
79	文化	*wénhuà*	culture	85	80	165	Empty	0.057521821
80	动作	*dòngzuò*	action	62	103	165	DE	0.000372138
81	作品	*zuòpǐn*	work	17	146	163	DE	0.000000000
82	任务	*rènwù*	task	13	150	163	DE	0.000000000
83	头发	*tóufǎ*	hair	34	127	161	DE	0.000000000
84	社会	*shèhuì*	society	44	115	159	DE	0.000000005
85	意义	*yìyì*	signifi-cance	11	147	158	DE	0.000000000

87	力量	*lìliàng*	strength	26	131	157	DE	0.000000000
88	眼神	*yǎnshén*	eyes	23	131	154	DE	0.000000000
89	车	*chē*	vehicle	44	109	153	DE	0.000000046
90	表面	*biǎomiàn*	surface	122	30	152	Empty	0.000000000
91	气	*qì*	gas	125	25	150	Empty	0.000000000
92	钱	*qián*	money	45	105	150	DE	0.000000309
93	价值	*jiàzhí*	value	30	118	148	DE	0.000000000
94	意见	*yìjiàn*	opinion	12	135	147	DE	0.000000000
95	经济	*jīngjì*	economy	62	83	145	DE	0.014532273
97	年纪	*niánjì*	age	108	34	142	Empty	0.000000000
98	功能	*gōng néng*	function	36	106	142	DE	0.000000001
99	家庭	*jiātíng*	family	32	110	142	DE	0.000000000
100	经验	*jīngyàn*	experience	16	126	142	DE	0.000000000

Table 3. Potential Alienable Possessives and the Occurrences of *De*. Rank over the 100 most frequent items, character, pinyin, gloss, frequency without 的 de (Empty) and with 的 *de* (DE), total frequencies, preference, and binomial *p*.

No	*Noun*	*Pinyin*	Gloss	Empty	DE	*n*	*Notes*	Prefer-ence	*binomial P*
1	手	*Shǒu*	hand	1010	1022	2032	Inalienable	DE	0.017082203
4	脸	*liǎn*	face	419	547	966	Inalienable	DE	0.000005240
6	党	*dǎng*	party	836	47	883	Politically inalienable	Empty	0.00000
7	妈	*mā*	mom	611	58	669	Inalienable	Empty	0.000000000
9	父亲	*fùqīn*	father	350	272	622	Inalienable	Empty	0.000239174
10	眼睛	*yǎnjīng*	eye	127	492	619	Inalienable	DE	0.000000000
12	妹	*mèi*	sister	602	5	607	Inalienable	Empty	0.000000000
15	母亲	*mǔqīn*	mother	310	224	534	Inalienable	Empty	0.000033342
16	身体	*shēntǐ*	body	201	307	508	Inalienable	DE	0.000000525
23	嘴	*zuǐ*	mouth	207	171	378	Inalienable	Empty	0.007400270
24	生命	*shēng-mìng*	life	156	208	364	Inalienable	DE	0.001016095
28	男人	*nánrén*	man	241	92	333	Inalienable	Empty	0.000000000

30	头	*tóu*	head	34	279	313	Inalienable	DE	0.000000000
31	孩子	*háizi*	child	56	255	311	Inalienable	DE	0.000000000
32	脚	*jiǎo*	foot	172	137	309	Inalienable	Empty	0.006261786
37	妈妈	*māmā*	mother	186	80	266	Inalienable	Empty	0.000000000
39	内心	*nèixīn*	heart	177	89	266	Inalienable	Empty	0.000000019
39	儿子	érzi	son	128	135	263	Inalienable	DE	0.044796180
40	父母	*fùmǔ*	parents	82	178	260	Inalienable	DE	0.000000001
46	全家	*quánjiā*	whole family	240	2	242	Inalienable	Empty	0.000000000
48	家	*jiā*	family	93	136	229	Inalienable	DE	0.000923784
49	兄弟	*xiōngdì*	brother	164	61	225	Inalienable	Empty	0.000000000
50	爸	*bà*	dad	216	8	224	Inalienable	Empty	0.000000000
55	爸爸	*bàba*	dad	162	50	212	Inalienable	Empty	0.000000000
56	全身	*quán-shēn*	whole body	181	25	206	Inalienable	Empty	0.000000000
60	爹	*diē*	father	184	16	200	Inalienable	Empty	0.000000000
61	女儿	*nǚ'ér*	daughter	102	98	200	Inalienable	Empty	0.054149674
64	眼	*yǎn*	eye	17	174	191	Inalienable	DE	0.000000000
70	双手	*shuāng-shǒu*	hands	109	70	179	Inalienable	Empty	0.000842170
71	娘	*niáng*	mother	165	13	178	Inalienable	Empty	0.000000000
76	耳	ěr	ear	146	22	168	Inalienable	Empty	0.000000000
96	姐妹	*jiěmèi*	sisters	127	15	142	Inalienable	Empty	0.000000000

Table 4. Potential inalienable Possessives and the Occurrences of *De*. Rank over the 100 most frequent items, character, pinyin, gloss, frequency without 的 de (Empty) and with 的 *de* (DE), total frequencies, additional notes, preference, and binomial *p*.

Glottodidattica

Language Teaching