The Information Packaging of the Do-Constructions in Chinese, Russian, and Czech

Chui, Kawai
Department of English
National Chengchi University
kawai@nccu.edu.tw

Yeh, Hsiang-lin
Department of Slavic
Languages and Literatures
National Chengchi University
verayeh@nccu.edu.tw

Lin, Melissa Shih-hui
Department of Slavic
Languages and Literatures
National Chengchi University
shihhui@nccu.edu.tw

Abstract

This study investigated the *do*-constructions in Chinese, Russian, and Czech, a predicate-argument structure comprised of the light verb 'to do' - *zuò* in Chinese, *delat'* in Russian, and *dělat* in Czech - and a verbal noun as the head in the accusative role, considering the linguistic traits and pragmatic use of the constructions in spoken and written discourse. The corpus results attested that the three languages not only have lexical and grammatical equivalences, they also demonstrate a functional equivalence in packaging information to define a type of action within the construction. Similar lexico-grammatical strategies are employed to encode tense and aspectual information of the predicates and various kinds of information about the nominal heads. The preference of the *do*-usage in the written genre is unequivocal in Chinese and Russian, suggesting that the structural change could have started as a writing style. The relative novelty of the *do*-usage to communicate generic or specific action events in Czech is evidence of language-specificity in pragmatic use.

1   Introduction

Light verb constructions, such as 'to take a turn', 'to give a rating', 'to make a second attempt', and 'to do a quick change', are cross-linguistic structures which have been studied in many languages including American English, British English, Irish English, Malaysian English, Zapotec, Spanish, Italian, Russian, Czech, Lithuania, Urdu, Persian, Japanese, Korean, and Chinese. The head nouns in the accusative position are 'verbal nouns'

and the constructions as a whole may be equivalent to the use of the head nouns as full-fledged verbs, also called 'heavy verbs', such as 'to turn' versus 'to take a turn', 'to rate' versus 'to give a rating', 'to attempt (to do something) for the second time' versus 'to make a second attempt', and 'to change quickly' versus 'to do a quick change'. A variety of issues were discussed from various approaches, among which the constructions were investigated in regard to argument or valency structures (Yim, 2020; Kettnerová, 2021; Kettnerová and Lopatková, 2020; Lin, 2014; Ronan and Schneider, 2017; Tadao, 2000). The corpus-based or descriptive approaches were employed to investigate the lexical, structural, semantic, and pragmatic properties of the light-verb constructions (Cuervo, 2010; Hernández, 2008; Huang and Lin, 2012; Huang et al., 2014; Maiko, 2020; Martínez Linares, 2013; Nolan, 2015; Ong and Rahim, 2021; Radimský, 2010; Ronan, 2014; Kovalevskaite et al, 2020), the linguistic distinction between the light verb and heavy verb usages (Beam de Azcona, 2017; Evteeva, 2017; Lu et al., 2020; Radimský, 2010; Tadao, 2000), the occurrences of the constructions in speaking and writing (Sundquist, 2020), the historical development of the constructions (Buckingham, 2014; Sundquist, 2018; Yim, 2020), the acquisition of the constructions in L2 contexts (Maiko, 2019; Sanromán Vilas, 2019), and the processing of light-verb structure (Wittenberg and Piñango, 2011). Little attention, however, has been paid to how the construction pairs the lexico-grammatical structure with meaning and function (Croft, 2014; Goldberg, 1995). As distinct from the usual predicate-argument combination where the event type is determined by the predicate, it is the

nominal argument of the *do*-construction that denotes a type of event in discourse.

This study investigates the light verb 'to do' forming a predicate-argument structure with a verbal noun as the head in the accusative role. This is called '*do*-construction' here. See the underlined parts in these English examples 'I <u>did some swimming</u> and headed home', 'I would <u>do less correcting</u> and more connecting', and 'We obviously need to <u>do a lot of praying</u>' from SKELL (skell_3_10 v1.8). The *do*-constructions in Mandarin Chinese ('Chinese' for short), Russian, and Czech are illustrated below. In Example 1, about psychological simplification, *zuò* 'to do' is the main verb and *jiǎnhuà* 'simplify' is the verbal noun as the head of the direct object which is quantified by *yīxiē* 'some' and characterized by *xīnlǐshàng* 'psychological'. In the Russian Example 2, 'to do author citations' is represented by the main verb *делать* 'to do' and the accusative form of the head noun *ссылки* 'citations' is qualified by *авторов* 'author'. In the Czech *do*-construction in Example 3, the verb *dělat* 'to do' and the accusative head noun *přehled* 'overview' as characterized by *dokonalý* 'perfect' together refer to the act of perfect overview.

(1) *nǐ   kěnéng   xūyào   **zuò**   yīxiē*
    2SG   may   need   do   some
    *xīnlǐshàng   de   **jiǎnhuà***
    psychological   DE   simplify
    'You may need to do some psychological simplification.'
(2) ***Делайте**   хотя бы   **ссылки**   на*
    do-IMP   at least   citation-PL-ACC   to
    авторов.
    authors
    'At least do author citations.'
(3) ***Udělali**   dokonalý   **přehled.***
    do-PST.PFV.3PL   perfect   overview
    'They did a perfect overview.'

The lexical and structural similarities of the *do*-constructions across Chinese as a Sino-Tibetan language, Russian as an East Slavic language, and Czech as a West Slavic language are not a coincidence. According to Natural Semantic Metalanguage (NSM), there are basic and universal semantic primitives that are conceptually simple and irreducible. "Evidence indicates that this highly constrained vocabulary and grammar

has equivalents in all or most languages of the world" (Goddard and Wierzbicka 2014:86). An inventory of semantic primes was proposed by Wierzbicka and colleagues as universal semantic fundamentals which been examined across a wide range of typologically different languages including Arrernte, Chinese, Ewe, French, German, Italian, Japanese, Lao, Malay, Mangaaba-Mbula, Maori, Polish, Russian, Spanish, and Yankunytjatjara (see the details in Goddard and Wierzbicka, 2014). 'DO' is a semantic primitive, and the direct lexical realization of this fundamental and universal concept of action is the *zuò*-verb in Chinese, the *delat'*-verb in Russian, and the *dělat*-verb in Czech. The lexical and grammatical behaviors of the three *do*-words as full-fledged verbs are not only identical, the verbs have also been undergoing a similar grammatical development and *do*-constructions are evolved. What remains obscure are the linguistic nature and the pragmatic use of the evolved structure in spoken and written discourse across the three languages.

The present study takes up the issue and asks how the basic semantic notion of DO engages in developing a widely-used structure. The lexical and structural equivalences of DO in Chinese, Russian, and Czech allow for cross-language investigation of the linguistic properties and the pragmatic use of *do*-constructions by carrying out a corpus analysis of *do*-cases derived from the major spoken and written genres. These research questions are addressed – What are the linguistic properties of the *do*-constructions in Chinese, Russian, and Czech? Are there genre differences between speaking and writing in regard to linguistic traits and occurrence rate? Is there language specificity in the pragmatic use of *do*-constructions? The corpus results enable establishment of a common functional construal of information packaging and discussion of the directionality of the structural spread and historical development of *do*-constructions.

2   The corpora and methods

Language use may vary between speaking and writing. For instance, the 3-word and 4-word lexical bundles predominated in spoken discourse, but a different combination of invariable function words and an intervening content word was prevalent in written academic discourse (Biber,

2009). The collocates for the verbs *have*, *make*, and *take* in conversation were also found distinct from those in informational writing (Conrad and Biber, 2009). The present study thus separates the spoken and written data for analysis. The data are drawn from the Corpus of Contemporary Taiwanese Mandarin 2017 (COCT), the Russian National Corpus, and the Czech National Corpus. First, the COCT documents written data from 1986 to 2017 in the areas of philosophy, religion, science, applied sciences, social sciences, history, geography, language, literature, arts, commerce, and recreation, totaling about 250-million words. The 2007-2014 spoken data consists of 6.6-million words from the sub-titles of Da Ai Journal, a TV program that documents inspiring stories of people and events around the world in areas of law, politics, finance, current events, science, living, fashion, culture, education, and arts. Second, the written data of the Main Corpus of the Russian National Corpus consists of 337-million words collected from fiction and news texts, and the spoken data, totaling 13.3-million words, are the recordings of public and spontaneous spoken Russian and the transcripts of the Russian movies. Data for this study are derived from 1981 to 2019. Last, the written corpus of the Czech National Corpus comprises 4255-million words collected between 1989 and 2014, and the spoken corpus consists of 7-million words produced in informal settings from 2002 to 2017.

The selection of data for this study met the criteria that the predicate of a clausal statement is the *do*-verb, namely *zuò* in Chinese, *делать (delat')* in Russian, and *dělat* in Czech, and the direct object comprises a verbal head noun which can be used as a full-fledged verb in other contexts. For instance, *jiǎnhuà* 'simplify' is the nominal head in Example 1, but the main verb in this statement of *wǒmen jiǎnhuà le jiàokēshū* 'We simplified textbooks.' Russian and Czech show the same usages, in that the accusative form *ссылки* in Example 2 is used as a verb in *Она ссылается на научные исследования* 'She cites scientific research', and the accusative head noun *přehled* 'overview' in Example 3 is a verb in *Z vrcholku hory lze přehlédnout široké okolí* 'From the top of the mountain you can overview the wide surroundings.' Upon this common lexical and grammatical foundation, cross-linguistic results are comparable. The search functions in the corpora

were used for data selection. The linguistic analysis of the selected data for each language was carried out by a first analyst and then checked and revised by a second analyst. Table 1 presents the sets of *do*-cases for the study. The Slavic languages consistently have a lower occurrence rate of *do*-cases than Chinese, whether in writing or in speaking. The overall frequencies in Chinese outnumber Russian by 6.6 times and Czech by 19.5 times, and the use of this construction is 3 times more prevalent in Russian than in Czech. The written cases predominate at 76.4% in Chinese and at 85.9% in Russian, whereas Czech shows a close distribution of the data across writing and speaking. These quantitative differences demonstrate that the semantic primitive of DO has been undergoing the same structural development across languages yet not in the same pace.

	Written	Spoken	Total
Chinese	15949	4925	20874
	76.4%	23.6%	100%
Russian	2725	448	3173
	85.9%	14.1%	100%
Czech	548	521	1069
	51.3%	48.7%	100%

Table 1: *Do*-cases in Chinese, Russian, and Czech.

3 The information packaging of *do*-constructions

The *do*-construction consists of two parts. The first part is the *do*-verb which means 'to act'; the second part is the noun phrase in the role of direct object. The nature of the action event is context dependent and determined by the encoding of information within the construction. The *do*-verbs in the Slavic languages are marked with tense and aspect information in all the cases. See the use of *делал* in Russian indicating the past and the imperfective aspect of the act of doing corrections in Example 4, and *udělal* in Czech showing the past perfective action of doing fake recordings in Example 5. Chinese, however, tends to encode these two types of information outside the construction by use of adverbials like *zuótiān* 'yesterday' and *míngtiān* 'tomorrow', *yǐjīng* 'already', *céngjīng* 'once', and *yīzhí* 'continuously'. The occurrence rate of aspect markers within the *do*-construction, like the perfective *le* as in

Example 6 about having done a very bad guide, the experiential *guò*, or the durative *zhe*, is low at 29.9% of the total 20874 cases.

(4) *Он* **делал** *бесконечные*
he do-PST-IPFV endless
исправления.
corrections
'He did endless corrections.'

(5) *Proč* **udělal** *ty* *falešné*
why do-PST.PFV.3SG these fake
zápisy *do* *svého* *deníku?*
recordings into own diary
'Why did he do these fake recordings into his own diary?'

(6) *wǒ* *céngjīng* *duì* *háizi* *de* *chuàngyì*
1SG once to children DE creativity
zuò **le** *yī* *gè* *hěn* *bù* *hǎo* *de*
do PRF one CL very NEG good DE
yǐndǎo
guide
'I once did a very bad guide to children's creativity.'

In the accusative position of the construction, the verbal noun as the head of direct object functions to represent an action event, and the noun phrase as a whole refers to a generic or specific event in discourse. A generic event refers to a general situation that is encoded by a bare nominal head without semantic characterization in the *do*-construction, such as *zuò chuànzhū* 'to do bead stringing' in Chinese (Example 7), *делать подтяжку* 'to do facelifting' in Russian (Example 8), and *udělali zátah* 'to do pulling' in Czech (Example 9).

Generic events – bare nouns
(7) *nóngfū* *xiàwǔ* *máng-wán* *nóngshì*
farmer afternoon work-finish farming
hòu *jiù* *huì* *zuò* *zài* *liángtíng* *shàng*
after then will sit at pavilion on
zuò **chuànzhū**
do bead
'The farmer, after finishing farming in the afternoon, would sit in the pavilion and do beading.'

(8) *A* *ваша* *жена* **делала**
and your wife do-PST.IPFV.3SG

подтяжку?
facelifting
'Did your wife do a facelifting?'

(9) **Udělali** **zátah**, *prohledali*
do-PST.PFV.3PL pull search-PST.PFV.3PL
a *našli* *spoustu* *zásob.*
and find-PST.PFV.3PL a lot of stock
'They did a pulling, searched and found a lot of stocks.'

A specific event, on the other hand, refers to a particular situation encoded with nominal qualification. Similar lexico-grammatical strategies, which are broadly categorized into definiteness, quantity, possession, and other qualifying properties, are employed to define specific events in Chinese, Russian, and Czech. See the following examples for the four types of strategies in the languages. First, definite referents of the *do*-events are marked by demonstrative words as in the Chinese 'do these three kinds of recycling' (Example 10), the Russian 'do such kind of recording' (Example 11), and the Czech 'do this discovery' (Example 12). Second, quantified referents are encoded by quantifiers or numerals such as 'do a little improvement' (Example 13), 'do one more stopping off' (Example 14), and 'do one adjustment' (Example 15). Third, the possessive information has to do with someone in possession of the nominal referents as in 'do our planting' (Example 16), 'do his own warnings (Example 17), and 'do my own smiling' (Example 18). Finally, other qualifying properties provide attributive information as in 'do a brief and seemingly meaningful pausing' (Example 19), 'do a witty literature review on nationalism' (Example 20), and 'do a significant smiling' (Example 21).

Specific events – definiteness
(10) *duì* *wǒ* *dàgài* *jīběnshàng* *huì* *zuò*
to 1SG probably basically will do
zhè *sān* *lèi* *de* *huíshōu*
this three kind DE recycle
'To me, basically, I probably will do these three kinds of recycling.'

(11) *Будучи* *на* *краю* *гибели* *ученый*
being on verge death-GEN scientist
делает *в* *своем*
do-PRS.IPFV.3SG in one's own
дневнике **такую** *запись: <...>.*
diary such recording

'Being on the verge of death, the scientist does such kind of recording in his diary.'

(12) *Když ona udělala* **tenhle**
when she do-PST.PFV.3SG this
objev a zavolala mi.
discovery and call-PST.PFV.3SG me
'When she did this discovery and called me.'

Specific events – quantity

(13) *jiāzhǎng hěn lèyì wèile háizǐ ānquán*
parent very happy for child safety
zuò **yīdiǎndiǎn** *gǎishàn*
do a little improve
'Parents are happy to do a little improvement for the safety of the child.'

(14) *Через полкилометра, на перекрестке –*
after half a kilometer at intersection
направо! Там делаем
to the right there do-PRS.IPFV.1PL
еще **один** *заход!*
more one stopping off
'After half a kilometer, at the intersection – to the right! We're doing one more stopping off there!'

(15) *Určitě tam udělám*
definitely there do-PRS.PFV.1SG
úpravu **jednu**.
adjustment one
'I will definitely do one adjustment there.'

Specific events – possession

(16) *wǒmen shì zài wúchénshì lǐmiàn*
we COP at dust-free room inside
zuò **wǒmen** *de zāizhòng*
do our DE planting
'We do our planting in a dust-free room.'

(17) *A то Министерство здравоохранения*
Otherwise Ministry health-GEN
обязательно делало бы
definitely do-PST.IPFV.3SG would
свои *предупреждения.*
one's own warning.PL
'Otherwise, the Ministry of Health would definitely do his own warnings.'

(18) *Udělal jsem* **svůj** *bolestný úsměv*
do-PST.PFV.1SG one's own painful smile
kolem úst.
around mouth
'I did my own painful smiling around my mouth.'

Specific events – qualifying properties

(19) *tā zuò le yī gè* **jiǎnduǎn ér shì**
he do PRF one CL brief and seem
yǒuyìhán *de tíngdùn*
meaningful DE pause
'He did a brief and seemingly meaningful pausing.'

(20) *Джон Бройи делает*
John Breuilly do-PRS.IPFV.3SG
остроумный *обзор* **литературы**
witty reviewing literature-GEN
о *национализме на глубину в*
about nationalism to depth at
четыре десятилетия.
four decades
'John Breuilly does a witty literature review on nationalism to the depth of four decades.'

(21) *Udělal jsem* **významný** *úsměv.*
do-PST.PFV.1SG significant smile
'I did a significant smiling.'

Most of the *do*-cases are specific action events, taking up 89% of the total in Chinese, 61.5% in Russian, and 66.4% in Czech. Differences are evident between writing and speaking. First, the frequency distribution of generic cases across the written and spoken data is about equal in Chinese and Czech, while Russian has the large majority of cases in the written texts. Second, the three languages align to show that specific events are the majority in writing, and the mean proportions are much higher in Chinese at 79.9% and Russian at 86.5% than in Czech at 56.3%. See Table 2.

Generic events	Written	Spoken	Total
Chinese	1103	1190	2293
	48.1%	51.9%	100%
Russian	1036	185	1221
	84.8%	15.2%	100%
Czech	148	211	359
	41.2%	58.8%	100%
Specific events			
Chinese	14846	3735	18581
	79.9%	20.1%	100%
Russian	1689	263	1952
	86.5%	13.5%	100%
Czech	400	310	710
	56.3%	43.7%	100%

Table 2: Frequency distribution of generic and specific *do*-cases in spoken and written data.

Considering the types of action events, among the 1453 types of verbal nouns in the written texts and 820 in the spoken texts in Chinese, 470 types of action are found in both writing and speaking. A larger variety of verbal nouns are used in the written mode at 67.7% than in the spoken mode at 42.7%, suggesting the vitality of the *do*-construction in written communication. The two Slavic languages have smaller sets of common types of action, a total of 85 in Russian and 40 in Czech. Like Chinese, Russian includes a lot more diverse types in writing at 70.8% than in speaking at 43%. Czech shows the opposite, in that there is a higher proportion of action types not found in the written texts at 72%. In regard to token frequencies, the shared action types crucially account for the large majority of cases in Chinese and Russian - 83.9% of all the Chinese written data and 86.7% of the spoken data; 80.8% of the Russian written data and 82.2% of the spoken data. In Czech, the highly repetitive use of the common types is seen only in the written cases at 89%. A large portion of the spoken cases, at 49%, demonstrate a much wider variety of action types in speech communication.

4   General discussion

Across Chinese, Russian, and Czech, the occurrence rates of *do*-constructions vary but the form, meaning, and function are equivalent. Croft (2014: 19) noted that "a construction (or any construction) in a language (or any language) used to express a particular combination of semantic structure and information packaging function." The *do*-construction comprises information that expresses a type of action which is denoted by the verbal noun rather than the *do*-verb, and the nature of the event has to do with the packaging of information about the accusative head in the context of use. The tense and aspectual information of the *do*-verb and the various kinds of information about the accusative head nouns together are essential to communicate a type of action event or a specific action event that is of interest in the context of use. This functional construal of the *do*-structure is evident in Chinese, Russian, and Czech, and the encoding strategies are cross-linguistically equivalent.

Across the written and the spoken texts, the structural and distributional analyses attested to the preferred communication of specific action events

in the *do*-constructions across languages. In terms of token frequencies, the two Slavic languages are similar to be less productive than Chinese; still, the occurrences across text genres are divergent between the two languages. The prevailing use of this grammatical structure in writing suggests that the encoding strategies have come to be adopted more readily as a writing style in Chinese and Russian. As to Czech, whether the *do*-construction tends to be a writing style or a speaking manner is not clear because of the relatively low occurrence rate. Regarding type frequencies, a verbal head being used in both the written and spoken texts was counted as a type. The Chinese data yielded a total of 470 verbal nominal heads, accounting for 83.9% of the cases in writing and 86.7% in speaking. Similar results are seen in Russian - 80.8% of the cases in writing and 82.2% in speaking refer to a set of 85 action events. Czech has a smaller set of 40 nominal heads that occurred in both types of text. Their occurrences account for 89% of the written data but only 51% of the spoken data. Taken the results together, the types of action that were brought up for discussion in the two types of discourse are considered to be more acceptable by language users and likely function as replicators that propagate the development of the *do*-structure. Synchronically, the cross-genre development is language-specific. The spread of *do*-usages was similar across written and spoken discourse yet only in Chinese and Russian. The Czech language manifests much novelty and diverseness in use of the *do*-construction in spoken communication.

Historically, the use of the *do*-construction in Chinese was far from common before the 20th century. From the Academia Sinica Ancient Chinese Corpus, 296 cases were derived, such as *zuò bùshī* 'to do almsgiving' in *The Water Margin: Outlaws of the Marsh* written in the late 14th century, *zuò gè zhèngjiàn* 'to do witnessing' in *Dream of the Red Chamber* in the 18th century, and *zuò jūtíng* 'to do short-time staying' in *The Scholars* in early 19th century. Further, a large portion of the cases at 48.6% were derived from the texts of *The Water Margin* which is known to be written in vernacular Chinese and considered as close to the spoken language. In a former study of the *do*-structure, the written data in the 20th century drawn from the 11-million-word Academia Sinica Balanced Corpus of Modern Chinese 4.0 (Sinica

Corpus) and the 382-million-word Chinese GigaWord 2 Corpus yielded a total of 3117 cases from 1981 to 2007 (Chui, 2018), and, in this study, five times more data from 1986 to 2017 were retrieved from the Corpus of Contemporary Taiwanese Mandarin. The diachronic data together reveal a contrast in the use of the *do*-construction before and after the 20th century.

In the Slavic languages, the historical Russian data in the Russian National Corpus are available from the 18th and the 19th centuries and a total of 2493 *delat'*-cases were drawn. 86.4% of the data were found in the texts between 1801 and 1900, such as *делать препятствие* 'to do hindering' (year 1775) and *делать великие описания* 'to do great descriptions' (1761-1765). Since the usage did not show a surge in the 20th century, 3173 cases in total, the Russian *delat'*-construction appears to be used earlier than Chinese in the 19th century, such as *делать наблюдения и открытия* 'to do observations and discoveries' (year 1867). The developmental tendency is similar in Czech. From the Czech National Corpus, a total of 134 *dělat*-cases were drawn from 1301 to 1900. Most of the data at 70.1% were from the texts in the 19th century, such as *dělat konec* 'to do an ending' (year 1894), after which the usage spreads gradually. See Table 3. In sum, the diachronic development of this grammatical structure since 1301 in Chinese and Czech supports Feltgen et al.'s (2017) claim that there could be a latency period of a change prior to the expansion of the use in the 19th or 20th centuries. In the present time, an S-curve for the development of the *do*-constructions is not seen due to the lack of a slow tailing off. It is also possible that the S-curve is not universal (Ghanbarnejad et al., 2014).

Chinese	Russian	Czech
1301-1900	1701-1900	1301-1900
N = 296	N = 2493	N = 134
1986-2017	1981-2019	1989-2017
N = 20874	N = 3173	N = 1069

Table 3: *Do*-cases in historical data.

Language change is initiated by language use (Feltgen et al., 2017). In the basic evolutionary model of language change, speakers replicate linguistic structures in utterances while interacting with other speakers, suggesting the usual directional spread from speech to writing (Blythe & Croft, 2012). The rise of the use of *do*-construction over the past 36 years (1981-2017) in Chinese and that of the *delat'*-construction in the past 38 years (1981-2019) in Russian demonstrate the predominant use in the written texts. The contemporary corpus data of these two languages further confirm the directionality of the structural change from writing to speaking as proposed in Chui (2018). In the literature of language change, cross-linguistic evidence was abundant to support the typical path of linguistic development from spoken to written discourse (Biber & Gray, 2011; Bybee & Hopper, 2001; Croft, 2000; Good, 2008; Hruschka et al., 2009). The reverse direction of change for the development of *do*-constructions is not common but by no means impossible. First, Biber & Gray's (2011) study attested that English complex noun phrases started in academic writing, but the structures did not spread to conversation. Second, if the change had started in the speaking environment, the use of the *do*- and the *delat'*-structures should have been more frequent in the spoken data. The statistics in Table 1 show the otherwise that the *do*-cases in writing are three times more frequent than those in speaking in Chinese, and the *delat'*-cases are six time more common in Russian. In Czech, the occurrence rates of the *dělat*-cases are about the same between the two genres, and, due to the relatively small amount of data, whether the language conforms to the common direction of linguistic spread that the spoken language affects the written language remains inconclusive.

Finally, the *do*-bare noun combinations are related to the use of the nominalized form as a full-fledged verb, in that both forms refer to a generic type of action, such as *zuò bǎguān* 'to do gatekeeping' versus *bǎguān* 'to gatekeep', *делать наколку* 'to do tattooing' versus *наколоть* 'to tattoo', and *dělat procházku* 'to do a walk' versus *procházet* 'to walk'. They are, however, different construals of the same experience. In the evolutionary framework for language change based on Hull's general analysis of selection for evolutionary systems (Blythe & Croft, 2012; Croft, 2000; Hull 1988, 2001), the canonical interactor in language change is the user who chooses what to say and how to say it. In other words, the *do*-usage and the full-verb usage could be the alternatives at the user's disposal in communication. It remains to

be seen whether the two usages engage in linguistic competition and whether the *do*-form bears any social or individual values as distinct from the full-verb usage.

In conclusion, the present study presented corpus evidence that DO not only has lexical and grammatical equivalences in Chinese, Russian, and Czech, this semantic primitive further demonstrates a less-known equivalence in structural change dated back to the fourteenth century in Chinese and Czech and to the eighteenth century in Russian. Along with the change is the evolvement of the pragmatic function of packaging information for defining a type of action within the *do*-constructions by virtue of common lexico-grammatical strategies. The preference of the *do*-usage in the written genre is unequivocal in Chinese and Russian, leading to our conjecture that the structural change could have started as a writing style. The relative novelty of the do-usage to communicate generic or specific action events in Czech is evidence of language-specificity in pragmatic use.

Acknowledgments

References

Beam de Azcona, Rosemary G. 2017. From the heavy to the light verb: An analysis of tomar 'to take'. *Lingvisticæ Investigationes*, 40(2):228–273.

Biber, Douglas and Bethany Gray. 2011. Grammar emerging in the noun phrase: The influence of written language use. *English Language and Linguistics*, 15:223–250.

Biber, Douglas. 2009. A corpus-driven approach to formulaic language: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3):275–311.

Blythe, Richard and Croft William. 2012. S-curves and the mechanisms of propagation in language change. *Language*, 88:269–304.

Buckingham, Louisa. 2014. Light verb constructions in the history of English. In Kristin Davidse, Caroline Gentens, Lobke Ghesquière and Lieven Vandelanotte, editors, *Corpus Interrogation and Grammatical Patterns*. John Benjamins, Amsterdam, pages 15–34.

Bybee, Joan and Paul J. Hopper. 2001. *Frequency and the Emergence of Linguistic Structure*. Oxford: John Benjamins.

Chui, Kawai. 2018. Directionality of change: Grammatical variation and Do-constructions in Taiwan Mandarin. *Concentric: Studies in Linguistics*, 44(1):65–88.

Conrad, Susan and Douglas Biber. 2009. *Real Grammar: A Corpus-Based Approach to English*. Pearson Longman, London, UK.

Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach.* London: Longman.

Croft, William. 2014. Comparing categories and constructions crosslinguistically (again): The diversity of ditransitives. *Linguistic Typology*, 18(3):533–551.

Croft, William. 2016. Comparative concepts and language-specific categories: Theory and practice. *Linguistic Typology*, 20(2):377–393.

Cuervo, Maria Cristina. 2010. Two types of (apparently) ditransitive light verb constructions. In Karlos Arregi, Zsuzsanna Fagyal, Silvina Montrul and Annie Tremblay, editors, *Romance Linguistics 2008: Interactions in Romance. Selected papers from the 38th Linguistic Symposium on Romance Languages (LSRL)*, Urbana-Champaign, April 2008. John Benjamins, Amsterdam, pages 139–156.

Evteeva, Mariya Yur'evna. 2017. On semantic opposition "VERBS vs TO MAKE + VERBAL NOUN". *Philology. Theory & Practice*, 10(76), Part 1:104–108.

Feltgen, Quentin, Benjamin Fagard, and Jean-Pierre Nadal. 2017. Frequency patterns of semantic change: Corpus-based evidence of a near-critical dynamics in language change. *Royal Society Open Science*, 4(11), 170830.

Ghanbarnejad, Fakhteh, Martin Gerlach, José M. Miotto and Eduardo G. Altmann. 2014. Extracting information from S-curves of language change. *J. R. Soc. Interface*, 11: 20141044.

Goddard, Cliff and Anna Wierzbicka. 2002. Semantic Primes and Universal Grammar. In Cliff Goddard and Anna Wierzbicka, editors, *Meaning and Universal Grammar: Theory and Empirical Findings, Volume I*, John Benjamins, Amsterdam, pages 41–85.

Goddard, Cliff. 2002. The search for the shared semantic core of all languages. In C. Goddard, & A.

Wierzbicka, editors, *Meaning and Universal Grammar - Theory and Empirical Findings, Volume I,* John Benjamins, Amsterdam, pages 5–40.

Goddard, Cliff and Anna Wierzbicka. 2007. Semantic primes and cultural scripts in language learning and intercultural communication. In Farzad Sharifian and Gary B. Palmer, editors, *Applied Cultural Linguistics: Implications for Second Language Learning and Intercultural Communication,* John Benjamins, Amsterdam, pages 105–124.

Goddard, Cliff and Anna Wierzbicka. 2014. Semantic fieldwork and lexical universals. *Studies in Language,* 38(1):80–127.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure.* University of Chicago Press, Chicago, IL.

Good, Jeff. 2008. *Linguistic Universals and Language Change.* Oxford: Oxford University Press.

Hernández, Roberto Mayoral. 2007. A variation study of verb types and subject position: Verbs of light and sound emission. In José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez and María José Cabrera, editors, *Romance Linguistics 2006: Selected papers from the 36th Linguistic Symposium on Romance Languages (LSRL),* New Brunswick, March-April 2006. John Benjamins, Amsterdam, pages 213–226.

Hernández, Roberto Mayoral. 2008. Mapping semantic spaces: A constructionist account of the "light verb" xordæn 'eat' in Persian. In Martine Vanhove, editor, *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations.* John Benjamins, Amsterdam, pages 139–161.

Hruschka, Daniel J., Morten H. Christiansen, Richard A. Blythe, William Croft, Paul Heggarty, Salikoko S. Mufwene, Janet B. Pierrehumbert and Shana Poplack. 2009. Building social cognitive models of language change. *Trends in Cognitive Science,* 13(11):464–469.

Huang, Chu-Ren and Jingxia Lin. 2012. The ordering of Mandarin Chinese light verbs. In Donghong Ji and Guozheng Xiao, editors, *Chinese Lexical Semantics: The 13th Chinese Lexical Semantics Workshop.* Springer, Heidelberg, pages 728–735.

Huang, Chu-Ren, Jingxia Lin, Menghan Jiang and Hongzhi Xu. 2014. Corpus-based study and identification of Mandarin Chinese light verb variations. In *Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects,* pages 1–10, Association for Computational Linguistics and Dublin City University. Dublin, Ireland.

Hull, David L. 1988. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science.* University of Chicago Press, Chicago, IL.

Hull, David L. 2001. *Science and Selection: Essays on Biological Evolution and the Philosophy of Science.* Cambridge University Press, Cambridge, UK.

Kettnerová, Václava and Markéta Lopatková. 2020. Reciprocity in Czech light verb constructions: The dependency perspective. *Jazykovedný Časopis,* 71(1):41–68.

Kettnerová, Václava. 2021. Optional valency complementations in Czech light verb constructions. *Linguistica Pragensia,* 31(1):7–27.

Kovalevskaite, Jolanta, Erika Rimkutė and Laura Vilkaitė-Lozdienė. 2020. Light verb constructions in Lithuanian: Identification and classification. *Studies about Languages,* 36:5–16.

Maiko, Tatsiana. 2019. Конструкции с опорным глаголом в речи изучающих русский язык как иностранный (Support (light) verb constructions in the speech of learners of Russian as a foreign language). In Iliyana Krapova, Svetlana Nistratova and Luisa Ruvoletto, editors, *Studi di linguistica slava. Nuove prospettive e metodologie di ricercar,* pages 285–301.

Maiko, Tatsiana. 2020. What can you give in Italian that you can't give in Russian? A contrastive study of constructions with the light verbs *dare* in Italian and *davat'/dat'* in Russian. In Joanna Szerszunowicz and Martyna Awier, editors, *Reproducible Multiword Expressions from a Theoretical and Empirical Perspective,* pages 33–54.

Martínez Linares, María Antonia. 2013. Light verb constructions in Latin American newspapers: Creative variants and coinages. *Spanish in Context,* 10(1):114–135.

Nolan, Brian. 2015. Determining light verb constructions in contemporary British and Irish English. *International Journal of Corpus Linguistics,* 20(3):326–354.

Ong, Christina Sook Beng. 2021. Nativised structural patterns of make light verb construction in Malaysian English. *Concentric: Studies in Linguistics,* 47(1):93–112.

Radimský, Jan. 2010. *Verbo-nominální predikát s kategoriálním slovesem (Verb-noun Predicates with a Light Verb).* Jihočeská univerzita, České Budějovice.

Ronan, Patricia and Gerold Schneider. 2017. Spanish infinitives borrowed into Zapotec light verb constructions. In Karen Dakin, Claudia Parodi and Natalie Operstein, editors, *Language Contact and Change in Mesoamerica and Beyond*. John Benjamins, Amsterdam, pages 56–80.

Ronan, Patricia. 2014. Complex predicates and light verb constructions in Modern Irish. *Revista Española de Lingüística Aplicada/Spanish, Journal of Applied Linguistics*, 27(1):140–167.

Sanromán Vilas, Begoña. 2019. Light verb constructions in spoken L2 English: An exploratory cross-sectional study. In Vaclav Brezina, Dana Gablasova and Tony McEnery, editors, *Corpus-based Approaches to Spoken L2 Production: Evidence from the Trinity Lancaster Corpus*. John Benjamins, Amsterdam, pages 181–206.

Sundquist, John D. 2020. Give as a light verb. *Functions of Language*, 27(3):280–306.

Tadao, Miyamoto. 2000. *The Light Verb Construction in Japanese: The Role of the Verbal Noun*. John Benjamins, Amsterdam.

Wittenberg, Eva and Maria Mercedes Piñango. 2011. Processing light verb constructions. *The Mental Lexicon*, 6(3):393–413.

Yim, Changguk. 2020. Productivity, richness, and diversity of light verb constructions in the history of American English. *Journal of Historical Linguistics*, 10(3):349–388.

Appendix. Abbreviations of linguistic terms.

CL	classifier
COP	copula
DE	morpheme *de*
GEN	genitive
IPFV	imperfective aspect
PFV	perfective aspect
PL	plural
PRF	perfective morpheme
PRS	present tense
PST	past tense
SG	singular
1	first person
2	second person
3	third person