跨語言空殼名詞平行語料庫建置

鍾曉芳[*]、謝承諭[**]、楊惟婷[***]、王淑琴[****]、林蒔慧[*****]、趙逢毅[******]、
顏郁哲[*******]

摘要[********]

「空殼名詞」－即意義相對空泛的名詞－其出現通常泛指非特定的想法、
問題等名詞，必須透過文章內容得知其指涉的意義為何。跨語言結構資料
庫為因應不同目的而建構的資料庫與日俱增，然現存的資料庫中，並無提
供空殼名詞的資源。本文章提供建置「空殼名詞跨語言結構資料庫」之內
容，其中包含五種語言（中文、日文、捷克文、德文以及越南語）的空殼
名詞跨語言對應。執行流程：1. 收集語料；2. 空殼名詞翻譯：以 Schmid
（1999；2000）的研究為本，將英語空殼名詞翻譯為中文、日文、捷克
文、德文以及越南語；3. 五種語言空殼名詞對照；4. 語料庫介面建立；
5. 建置「空殼名詞跨語言結構資料庫」完成。語言學研究需要共享的語
言資料，如各類詞表。這些詞表的共享激勵了相關研究能量，其共通點是
可以利用電腦程式處理，並可成為人工智能或數位學習程式的應用資源。
本研究所提供的提供人工標註翻譯及檢索平台可作為未來更多空殼名詞
或跨語種研究的資料，應用於人工智慧或數位學習資源。

關鍵詞：空殼名詞、跨語言、資料庫、翻譯、平行語料庫

_____

* 國立政治大學英國語文學系教授
**國家教育研究院語文教育及編譯研究中心助理研究員
***國立政治大學華語文教學碩士學位學程碩士畢業生
****國立政治大學日本語文學系教授
*****國立政治大學斯拉夫語文學系教授

The Construction of the NCCU Shell Nouns Cross-Language Parallel Corpus (NCCU-SNPC)

Siaw-Fong Chung[*], Chen-Yu Chester Hsieh[**], Wei-Ting Yang[***], Shu-Ching Wang[****], Shih-Hui Lin[*****], August F.Y. Chao[******], Yu-Che Yen[*******]

Abstract[********]

The concept of a shell noun can be understood as a noun having a bare meaning to express an uncertain concept. This involves words such as *idea*, *problem*, *issue*, and some other general nouns. Normally, most shell nouns are highly contextualized. Their meanings are largely based on the context where they appear. Their existence in other languages, however, is overlooked. At present, cross-language structured databases have been constructed to meet various purposes, and they continue to grow. Yet, among current corpora, we rarely see a multiple-language translation of shell nouns. For this purpose, it is worthwhile to conduct a database with example sentences that can compare their uses in various languages. The aim of this project is to create one such database – the NCCU Shell Nouns Cross-Language Parallel Corpus (NCCU-SNPC), which includes data from Chinese, Japanese, Czech, German, and Vietnamese. These are the steps: (a) We first collected related language data from Schmid (1999, 2000). (b) The language data was translated into the target languages, based on the list and sentences we collected. (c) The mapping of different languages was carried out. (d) A platform containing the five languages with searchable function was created. (e) The final step was to finalize the construction of the NCCU-SNPC and to check for consistency. Linguistic studies require shared data of different languages, and it has become a trend to share data to inspire related research. Using our data, more research can be carried out so as to provide human annotated data for applications in A.I. or digital learning resources.

Key words: Shell nouns, cross-language study, database, translation, parallel corpus

1.　Introduction

Shell nouns "turn physically observable events as well as abstract relations into mentally perceivable IDEAS" (following Lakoff & Johnson, 1980, concepts are written in small capitals), and "[c]onceptual shell nouns epitomize the image of conceptual shells which contain mental entities" (Schmid 2000 188). Table 1 shows examples of Schmid's shell nouns.

Table 1: Examples of shell nouns (Schmid 2000 4)

Class	Examples
Factual	*fact, thing, point, problem, reason, difference, upshot, result*
Linguistic	*news, message, rumor, report, order proposal, question*
Mental	*idea, motion, belief, assumption, aim, plan, decision*
Modal	*possibility, truth, permission, obligation, need, ability*
Eventive	*act, move, measure, reaction, attempt, tradition, trick*
Circumstantial	*situation, context, place, area, time, way, approach*

The following are some examples from Schmid (2000 188):

(1) (a) **the thought** *that that man [Charles] will one day be King*

　 (b) **the old theory** *that you can tell a woman's true age by looking at her hands*

　 (c) **the idea** *of the socialist state*

Many shell nouns often "do not only **create** concepts, but they also **mean** *concepts*" (Schmid 2000 188). Shell nouns have been widely researched in English as their uses in other languages are few (Aktas & Cortes 2008 on writing experiences; Flowerdew 2003 on stance making; Francis 1986 on anaphoric nouns). Most studies on shell nouns have focused on academic writing (e.g., natural sciences, social sciences, humanities, etc.; cf. Flowerdew & Forest 2017). Table 2 shows the shell nouns that mean IDEA.

Table 2: Shell nouns for IDEA

Shared Semantic Features	[MENTAL]; [CONCEPTUAL]
Frame	Mental; focus on the propositional content of IDEA
Nouns	*point, idea, position, issue, theory, notion, thought, principle, rule, subject, image, myth, law, theme, concept, secret, scenario, wisdom, hypothesis, thesis, logic, topic, doctrine, teaching, maxim, stereotype, rationale, metaphor, axiom, dogma, credo, motto, precept*

According to Schmid (2000 188-89), these IDEA nouns portray the "PSYCHOLOGICAL STATES imputed to the EXPERIENCERS by the speaker." Sometimes, these ideas represent a thought, a fact, an event, etc. As mentioned above, Schmid's (2000) aim was to find out what writers were referring to by using these shell nouns; our goal was to find out how these shell nouns could be translated into other languages.

Research on languages other than English usually starts with

Schmid's (1999 & 2000) list of shell nouns (e.g., Hsieh, C-Y. C. 2012; Aktas, R. N., & Cortes, V. 2008). Based on this notion, this paper also aims to provide a list of shell nouns in other languages for comparison with the English ones. Our study aimed to provide the translation of not just the list of shell nouns but also the target expressions in which they appear to understand the functions and features of shell nouns when they appear in context. When we translated these target expressions into different languages, we saw how they were used. This study intended to introduce a newly developed cross-language database of shell nouns as a tool for researchers interested in examining how shell nouns are used in spoken and written discourse.

The aim of constructing the NCCU Shell Nouns Cross-Language Parallel Corpus (NCCU-SNPC) was to show the diverse language features of shell nouns. In this study, the aim of constructing a shell noun cross-language database was to collect and present shell nouns in different languages, including Chinese, Japanese, Czech, German, and Vietnamese, mostly receiving little attention in studies of shell nouns. These languages demonstrate unique linguistic features. Therefore, this cross-language parallel corpus intends to include five different languages from various language families, including Sino-Tibetan languages, Japanese-Ryukyuan languages, Slavonic languages, Indo-European languages, and Austroasiatic languages, to explore how shell nouns can exist and be translated among languages. To understand how the collected data from all these languages were managed, this paper will demonstrate the data collection procedure. The inter-translation among the five languages was conducted by scholars and students from different departments and fields. A description of how the interface was created will be introduced to understand how the shell noun cross-language database was built and how it can be used as a valuable tool to expand the scope of research on shell nouns.

1. Literature Review

This section will present the basics of Schmid's (2000) shell noun framework and review previous research on shell nouns in English and other languages.

2.1 Shell Noun Basics in English

The content of shell nouns is generally unspecified, but Schmid (2000) found that shell nouns can link to and encapsulate chunks of propositional information (e.g., events and abstract relations) in co-texts. Distinct from other types of nouns, shell nouns must, and can only, be interpreted together with their shell content. In order for this group of abstract nouns to connect to the shell content, they tend to be used in particular lexicogrammatical patterns, as "shell-noun constructions." According to Schmid (2000), in English, shell-noun constructions primarily include four groups of lexicogrammatical patterns: N-Cl, N-*be*-Cl, *th*-N, and *th*-*be*-N, as illustrated in (2). As can be seen, these constructions mostly involve linguistic devices, such as demonstratives and complementizers, that help link the abstract noun to another piece of (propositional) information in the co-text. The emphasis on the grammatical constructions with which abstract nouns co-occur and the functions that ensue differentiates Schmid's (2000) framework from other similar ones such as those for general nouns (Halliday & Hasan 1976) and anaphoric nouns (Francis 1986). Therefore, it is interesting and relevant to conduct cross-linguistic research, because typological differences such as word order, case-marking, and clause-marking can give rise to distinctive patterns in form and function (e.g., Hsieh 2020).

(2) Examples of shell-noun constructions from Schmid (2000 22) (The words in bold are shell nouns and the underlined part is the shell content and the shell-noun construction):

   (a) N-Cl: *Mr. Bush said Iraq's leaders had to face* <u>the</u> **fact** <u>that the rest of the world was against them.</u>

   (b) N-*be*-Cl: <u>The</u> **advantage** <u>is that there is a huge audience that can hear other things you may have to say</u>.

   (c) *th*-N: *(<u>Mr Ash was in the clearest possible terms labelling my clients as anti-semitic.</u>) I hope it is unnecessary to say that <u>this</u> **accusation** is also completely unjustified.*

(d) *th-be*-N: (I won the freshmen's cross-country. – *Mm.*) <u>*That was*</u> *a great* **achievement** *wasn't it?*

Schmid (2000) further classified shell nouns, based on their semantic features, into six categories, including factual, linguistic, mental, modal, eventive, and circumstantial uses, as shown in Table 1 previously. In each of the six uses, the shell nouns were further grouped into smaller sub-categories according to the semantic "frame" that the nouns belonged to collectively. For example, nouns that expressed causal relations such as "reason" and "result" were subsumed under the "causal uses" sub-category, which in turn was one of the sub-types of "factual uses". Another noteworthy characteristic of this semantic taxonomy is the classification of the sense, rather than the form, of shell nouns. As a result, the same form, such as 'point', can occur in more than one category (in this case, factual, linguistic, and mental uses) due to its polysemy in English. Based on this classification system, Schmid (2000) and a number of later studies also discussed and displayed the correlation between the semantic categories and the co-occurring patterns (Vergaro 2018; Vergaro & Schmid 2017).

In addition to their distinctive characteristics in syntax and semantics, shell nouns are also remarkable in their functions on different levels. Having examined the corpus data of more than 700 shell nouns, Schmid (2000) suggested that this huge group of abstract nouns could help achieve a number of semantic, textual, and cognitive purposes. Semantically, shell nouns "characterize" the proposition to which they co-refer in the text, while pragmatically they help link the messages and manage the topic and focus of the discourse. On the cognitive level, which was most important for Schmid (2000), shell nouns enable the writer/speaker to partition a concept, present the concept temporarily as a thing-like entity that can be manipulated, and integrate the complex information encoded in the abstract and the proposition.

2.2 Applied Linguistics Studies on Shell Nouns in English

The framework of shell nouns has also been widely adopted by applied linguists of English to analyze various types of discourse. One of the most examined areas has been academic discourse. In addition to identifying the functions of shell nouns, such as discourse organization and stance-marking (e.g., Charles 2007; Flowerdew 2003; Swales 2001), this line of research has further pinpointed a number of factors that may influence the distribution of shell nouns in academic texts and talks, including register/mode (speech v. writing) (Benitez-Castro 2014; Fang & Dong 2021), genres (e.g., lectures, journals, and textbooks; Flowerdew & Forest 2015; Kou 2022; Liu & Deng 2017), academic disciplines (Charles 2007; Dong, Fang, & Qui 2020), first language (Dong & Fang 2021; Flowerdew 2006, 2010; Tahara 2017), and writing experience (experts v. students) (Aktas & Cortes, 2008). Shell noun studies that drew on learner data showed that L2 student writers, compared with L1 student writers and expert counterparts, either produced fewer types of shell nouns (Flowerdew 2006, 2010) or used shell nouns in distinctive patterns and for different purposes (Aktas & Cortes 2008). Pedagogical implications can thus be drawn from these research findings (Flowerdew 2006 & 2010).

Researchers have also, albeit sporadically, looked into the use of shell nouns in legal and journalistic discourse. For example, Mazzi (2005) compared the use of "reasons" and "grounds" in judicial texts as argumentative keywords and found that these words, while semantically close, tended to co-occur with different lexicogrammatical patterns to serve different persuasive functions in judicial argumentation. Focusing on texts of judicial argumentation, Mazzi (2010) suggested that constructions such as "this/these/that/those + labeling noun" were linguistic resources for United States Supreme Court Justices when taking a stance and organizing their discourse in composing judgments.

Last but not least, the use of shell nouns in journalism has also been discussed. For example, Fang and Dong (2021 232), based on the data extracted from the ICE-GB corpus of contemporary British English, argued that shell nouns could function as discursive devices for formal and persuasive texts, such as news reports, that allowed the writer to organize the journalistic text while "blurring the distance in terms of formal role relationship between the news media and the public as well as the

separation between the institutional nature of news reporting and personal belief."

2.3 Shell Nouns in Other Languages

Although the notion of shell nouns was first examined and predominantly studied in English, similar uses of abstract nouns and constructions have also been attested and analyzed in other Indo-European languages, including Catalan (Cuenca 2014), French (Roze et al. 2014), German (Hundt & Oppliger 2022), and Spanish (Carretero 2016; Curry & Pérez-Paredes 2021). Some of the studies focused on the method for identifying and annotating shell nouns in a particular language (e.g., Roze et al. 2014), while others investigated the use of specific noun-based patterns in a given context (e.g., Cuenca 2014).

One major line of research compared the use of shell nouns in English and another European language in discourse. For example, Cuenca (2014) analyzed the discourse of the Catalan Parliament and pointed out that the combination of a demonstrative and an abstract noun was often employed by politicians to encapsulate information and organize political debates. Examining the "stance noun + that/de que" construction in English and Spanish academic blogs with themes related to COVID-19, Curry and Pérez-Paredes (2021) concluded that how the construction was employed to express the stance noun showed how the pandemic was framed in each language.

Comparing the "N-is/ist" constructions in English and German, Hundt and Oppliger (2022) found that while article use in German constructions was largely determined by the abstract noun, that in English was influenced by different variables. Hundt and Oppliger (2022) further argued that as it was easier to achieve a formal syntactic separation in English, English N-is constructions were more likely to be used as focalizers than their German counterparts.

In a study on the use of shell nouns in Czech, Lin (2023) affirmed that employing the instrumental case with shell nouns is a distinctive linguistic feature specific to the Czech language. However, Lin and Šebesta (2023) demonstrated that this case choice was significantly less common in learner corpus data, particularly in texts generated by Czech learners

whose native language was Chinese.

By comparison, the topic of shell nouns has attracted much less attention from researchers of typologically more distant languages, such as Japanese and Korean, although some studies have in fact reported similar usage patterns of abstract nouns (e.g., Ahn & Yap 2020; Fujii 2000; Kim 2011; Takara 2012). Among the very few linguists of Asian languages that have paid special attention to shell nouns, Hsieh (2012, 2018, & 2020) has conducted a series of studies on the use of individual shell nouns and shell-noun-based expressions in Mandarin Chinese. Hsieh (2012) examined the use of expressions that contained Chinese shell nouns, such as wentishi 'the problem is' and shishishang 'in fact,' in spoken discourse, arguing that these constructions enabled Mandarin speakers to negotiate their stance in interactions in different ways. Furthermore, Hsieh (2018) pointed out the role of wentishi, a shell-noun-based expression, in stance-taking and turn-taking, whereas Hsieh (2020) indicated that the Chinese shell noun yisi 'meaning, idea, intention' was recurrently used in conversational repair and that yisi displayed characteristics specific to Chinese in its co-occurring patterns and pragmatic functions.

A more recent line of research examined the use of Chinese shell nouns in different genres (Kou 2022; Yang 2022). For instance, in Kuo's (2022) study comparing the frequencies, semantic types and syntactic constructions of shell nouns in Chinese popular and professional science articles, Kou found that while shell nouns are more frequently employed to express explicit evaluation in popular science articles, those utilized in the professional genre tend to be more informative, objective or indicative of scientific rationality. Another study by Yang (2022) focused on e-petition posts in Taiwan and observed that this genre in general, and each rhetorical move that constituted this genre in particular, showed a preference for specific shell nouns and shell-noun constructions, partly due to their respective semantic and pragmatic features.

As shown in the literature review above, research on shell nouns in different languages and contexts is promising and is still on the rise. An open-access, cross-linguistic resource with a special focus on shell nouns, such as the multilingual parallel corpus that is about to be introduced in the following, is thus relevant and vital.

3. Steps in Creating the NCCU Shell Nouns Cross-Language Parallel Corpus (NCCU-SNPC)

The NCCU-SNPC is a multilingual parallel corpus that provides cross-language information retrieval. It was designed to consider factors such as translation error, language proficiency, grammar comparison, and so on. Detailed information about its creation is shown in Figure 1 and will be introduced in the following sections.
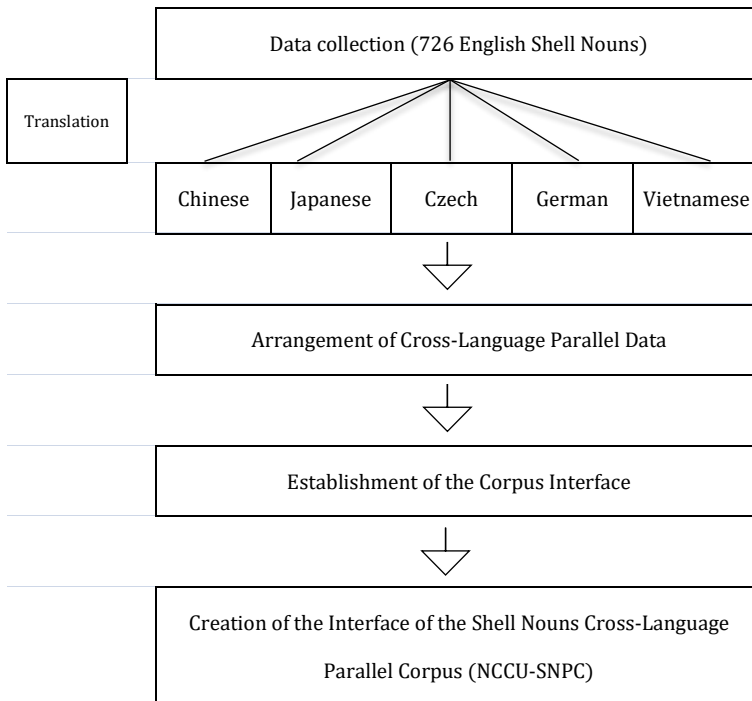
Figure 1. Data collection and digitalization procedures

3. Methodology

3.1 Data Collection

Step 1: Collection of English shell nouns proposed by Schmid (2000)

Schmid (2000) pointed out that shell nouns are abstract, unspecific, and can only be interpreted together with the shell content. In his book *English Abstract Nouns as Conceptual Shells*, Schmid proposed a total of 726 shell nouns, such as "fact," "problem," and so on. Moreover, Schmid classified these nouns into six categories based on their semantic functions (as shown in Table 1.)　In this study, 726 English shell nouns were provided but only 148 target expressions with complete sentences were used to be included as target expressions. These target expressions are further classified based on their category, and then saved as an Excel file shown in Figure 2. Column A (far left) shows the semantic categories of the shell nouns; Column B displays the sub-categories of the semantic functions; Column C lists the English shell nouns; and Column D presents the target expressions.

A	B	C	D	E	F	G	H	I	J
factual uses	neutral uses	thing	You're 61 now and it's time you settled down.　"the thing is that he needs a lot of loving. (p.94)						
factual uses	neutral uses	fact	We never expected John that er they would just change it like that in the light of the fact that they						
factual uses	neutral uses	phenomenon	"[...].Later, mummy herself becomes the prototype for her aspirations and a little envy begins to						
factual uses	neutral uses	case							
factual uses	neutral uses	business							
factual uses	neutral uses	point							
factual uses	causal uses	reason	But Stehli has a good reason to write her story. (p103)						
factual uses	causal uses	cause							
factual uses	causal uses	ground							
factual uses	causal uses	result	When women are young, their careers and pension rights would be damaged if they took time of						
factual uses	causal uses	consequence							
factual uses	causal uses	impact							
factual uses	causal uses	payoff							
factual uses	causal uses	sequel							
factual uses	causal uses	outcome							
factual uses	causal uses	effect							
factual uses	causal uses	upshot							

Figure 2. Collection of 726 English shell nouns and 148 target expressions

Step 2: Translation data from English to Chinese, Japanese, Czech, German, and Vietnamese

To observe and compare the linguistic features of shell nouns in cross-languages, we invited professors and graduate students from different departments, including teaching Chinese as a second language, Department of Japanese, Slavic languages and literature, and Department of English, to translate English shell nouns and the target expressions from Schmid (2000) into the target languages. The results of these translated shell nouns into the target languages were provided by our authors who are experts in respective languages, and most authors checked with the native speakers when they had doubts. Translated tools were used to make sure the meanings of the shell nouns were carefully translated without misleading the original context. Any disagreements were discussed until consensus was reached.

3.1.1 Translation data from English to Chinese

For Chinese, we used the Chinese-English thesaurus system online resource developed by the National Academy for Educational Research (https://coct.naer.edu.tw/bc/) to translate shell nouns from English to Chinese, because this bilingual query system shows word collocations and bilingual target expressions at the same time. Figure 3 presents the steps used in this online system for translation. First, the English shell noun 'thing' was typed into the search bar. Next, the upper-right block showed several possible translations. Then, the top two or three high-frequency translations were chosen (in this example, the translation of 'thing' was 'dongxi', '*shi*', and '*shiqing*'.

Figure 3. Using the Chinese-English thesaurus system for Chinese translation

Because the language structure of English and Chinese is different, we needed a Chinese grammar handbook to guide us in how to translate properly. According to *Chinese: A Comprehensive Grammar*, the author mentioned that Chinese modifiers such as clauses are placed before nouns while in English they follow nouns, as shown in (3a) and (3b), respectively:

(3) (a) English: "…the government side would not agree to the demonstrator's *demand* that cameras be allowed into the meeting to film the event."

(b) Chinese: "政府方不應該允許示威者攜帶照相機進入會議中拍攝的要求。"

In the original sentence, the clause is followed by the shell noun '*demand*', while in the Chinese translation the shell noun '要求' (in red font) is at the end of the sentence.

B	C	D	E	F
		中文翻譯	英文例句	中文翻譯
neutral uses	thing	東西、問題	You're 61 now and it's time you settled down. "the thing is that he needs a lot of loving.	你現在 61 歲，是時候談安定下來了，「問題是，他需要非常多的愛。」
neutral uses	fact	其實、事實（上）	We never expected John that er they would just change it like that in the light of the fact that they're going to change local government across the country anyway.	我們從來沒想過約翰會那樣改變它，事實上，無論如何他都會改變全國各地的地方政府。
neutral uses	phenomenon	現象	"[...].Later, mummy herself becomes the prototype for her aspirations and a little envy begins to creeps in." It is this supposed envy of mummy that is the most dangerous aspect of the new phenomenon.	之後，木乃伊成為了她願望中的原型，同時嫉妒感一點一滴的增長。「木乃伊的這種嫉妒之心正是這種新出現的現象中最危險的部分」

Figure 4. Shell nouns in the Chinese-English translations

Translating English target expressions from English to Chinese by using the Chinese-English thesaurus system, this study input these translated shell nouns, highlighted in red, into Column D of the Excel file, shown in Figure 4. While these translated shell nouns demonstrated their semantically unspecific and abstract nature, contextual meaning is one of the factors influencing how the target expressions were translated. Considering this factor, we found some target expressions in Schmid's book and further translated them to the target language expressions (see Table 3 for the rough translations; Reference: COCT NAER https://coct.naer.edu.tw/bc/).

Table 3: Shell nouns translated into Chinese

空殼名詞類別	例子
事實詞彙	事實(上)[1]、東西、事物、難題、困境、理由、原因、差距、差異、落差、結果、成果
語言詞彙	消息、訊息、傳聞、傳言、流言、報導、報告、提案、方案、問題、疑問
心理詞彙	概念、構想、動作、信念、理念、假設、預設、目的、目標、構想、藍圖、抉擇、選擇
情態詞彙	可能性、可行性、許可、義務、責任、需求、所需、能力、實力
事件詞彙	行動、行為、走向、動作、措施、手段、反應、回應、意圖、慣例、訣竅、伎倆
狀況詞彙	情況、處境、脈絡、背景、領域、時候、方式、方法、某事、作法

3.1.2 Translation data from English to Japanese

First, we used online translation software, such as Google and DeepL, to translate the English shell nouns into Japanese, and then confirmed and

---

[1] In Schmid's (2000) study, the list of shell nouns include 'fact' as in Chinese often appears to be '*shishishang* (事實上)' in the Chinese-English thesaurus system online resource developed by the National Academy for Educational Research (https://coct.naer.edu.tw/bc/). This is because the inputted keywords will be translated according to the sentences found in the interface. In most cases, 'fact' or 'in fact' will be translated as '*shishishang* (事實上)'. Therefore, this study decided to make the English shell noun 'fact' into the Chinese phrase '*shishishang*.'

corrected the Japanese presentation to ensure that the translations were consistent with the original texts. For the selection of vocabulary, we referred to online dictionaries such as Eizirou on the Web (https://eow.alc.co.jp/), Weblio (https://ejje.weblio.jp/), and Goo Dictionary (https://dictionary.goo.ne.jp/). To present a clear correspondence between the English and Japanese shell nouns, the following steps were followed: (A) select the corresponding Japanese words according to the context; (B) translate the nouns without affecting the naturalness of the Japanese expressions. Regarding (A), for example, the word "proof" in the original text in (4a) was translated as '證明' by the online translation software, as shown in (4b), but in this context, 'proof' referred to evidence of something, so it was more appropriate to use '証拠' or '証し', so the translation of (4c) was revised and the corresponding Japanese word was treated as the shell noun.

(4) (a) "That would be the final proof that everybody, including us, has gone nuts."

(b) "それは、私たちを含め、みんながおかしくなってしまったことの最終的な証明になる。"

(c) "それは、私たちを含むすべての人が気が狂ったという最終的な証拠になるだろう。"

　　Regarding step (B), for example, the shell noun 'difference' in the original text in (5a) was translated as the verb '違う' by the online translation software, such as in (5b). However, to clearly present the correspondence between the English and Japanese shell nouns, the noun form of the verb '違い'　was used for the translation, as shown in (5c).

(5) (a) "Of course, the fundamental difference between Radio One and other radio stations is that it's a public service broadcaster."

(b) "もちろん、ラジオ・ワンが他のラジオ局と根本的に違うのは、それが公共放送であることだ。"

(c) "もちろん、ラジオワンと他のラジオ局の根本的な<u>違い</u>は、公共放送局であるということである。"

However, there were some shell nouns that could not be translated into corresponding nouns in Japanese and instead had to be translated into verb or adverbial sentences, which shows the difference in language structure between English and Japanese. For example, in the original text in (6a), although there is the corresponding noun '談話' for '<u>talk</u>' in Japanese, it would have been very unnatural to translate the original text in (6b) as '〜という談話があった,' so the verb sentence '〜と言われていた' was used for the translation. The target expressions that could not be translated into nouns in Japanese accounted for a certain amount of the target expressions in the translations. This implied that the use of shell nouns in English was broader than in Japanese.

(6) (a) "After Lukoil, Russia's largest state-owned oil-producer, had joined the project earlier this year, there was <u>talk</u> that a contract to develop the oilfields would soon be agreed upon."

(b) "今年の初めにロシア最大の国営石油会社であるルコイユがこのプロジェクトに参加したことで、近々、油田開発の契約が結ばれるのではないかと<u>言われていた</u>。"

B	C	D 日文翻譯	E 英文例句	F 日文例句
neutral uses	thing	こと	You're 61 now and it's time you settled down. "the thing is that he needs a lot of loving.	もう61歳だし、そろそろ落ち着いてもいいんじゃない？"彼にはたくさんの愛情が必要だということである。
neutral uses	fact	事実	We never expected John that er they would just change it like that in the light of the fact that they're going to change local government across the country anyway.	全国の地方自治体を変えようとする事実に照らして考えると、ジョンがそのように変えることを期待したことがなかった。
neutral uses	phenomenon	現象	"[...].Later, mummy herself becomes the prototype for her aspirations and a little envy begins to creeps in." It is this supposed envy of mummy that is the most dangerous aspect of the new phenomenon.	その後、ミイラ自身が彼女の願望の原型となり、少しの羨望が忍び寄ってくる。"このミイラへの羨望こそが、この新しい現象の最も危険な側面なのである。
neutral uses	case			
neutral uses	business			
neutral uses	point			
causal uses	reason	理由	But Stehli has a good reason to write her story.	しかし、Stehliには彼女の物語を書く正当な理由がある。
causal uses	cause			
causal uses	ground			
causal uses	result	結果	When women are young, their careers and pension rights would be damaged if they took time off for their children, so grandmothers must help. The result is that you often know	女性が若いとき、子供のために休暇を取ると、キャリアと年金が損なわれるので、祖母は助けなければなりません。その結果、しばしば自分の子供よりも孫のこ

Figure 5. Shell nouns in the Japanese-English translations

3.1.3 Translation data from English to Czech

There were three steps to translating the English data into the Czech language. First, based on the English target expressions with shell nouns in Schmid (2000), we tried to find equivalent translations into Czech. Instead of word for word, the translation task focused more on the target expressions. However, in the cross-language comparisons, we had to take into consideration the fact that there might be significant differences in the frequency of occurrence of these shell nouns in the compared languages. In other words, it was possible to have translation equivalents, which then were necessarily considered from the point of view of their frequency. According to Kolhatkar and Hirst (2014 499), shell nouns occur frequently in all kinds of texts, and belong to nouns that are frequently used. However, in the process of translation, it is possible that the consequent status of expressions that in one language belonged to high frequency nouns, while in the target language the corresponding translation of lexical item belonged to infrequent one. Therefore, the second step was to check the translations with the help of Treq[2] and InterCorp[3] in the Czech National

---

[2] Treq: https://treq.korpus.cz/

[3] InterCorp: https://intercorp.korpus.cz/

Corpus. Treq is used to look up translation equivalents in bidirectional Czech-foreign language dictionaries automatically extracted from parallel texts in InterCorp, in which each text has a Czech counterpart, either original or a translation, aligned with one or more foreign-language versions (Rosen, Vavřín, & Zasina 2018). Figure 6 below presents a screenshot of how to use these tools to check the translations. In this example, the English shell noun "reason" was typed as a query. The block on the left shows several translation equivalents and their frequency via Treq, and the bigger block on the right shows the parallel texts in which "reason" and its Czech equivalents are used. Afterwards, we checked whether the Czech translations achieved in the first step matched the most or the second-most frequent translations via Treq and InterCorp, and in this case the translation was confirmed to be "*důvod*."

Figure 6. Shell nouns in the Czech-English translations

However, if they did not match, the third step was to ask for a third opinion to figure out whether the top two or three high-frequency translations via Treq and InterCorp could replace the original translation, or whether it was better to maintain the original translation. The most important and critical consideration depended on the context, which

tended to influence the lexical meaning of the shell nouns. After translating the data into Czech, the data were placed into Column F of the Excel file shown in Figure 7. The linguistic items that were assumed to function as shell nouns in Czech are highlighted in green.

B	C	D	E	F
		捷克文翻譯	英文例句	捷克文例句
neutral uses	thing	to	You're 61 now and it's time you settled down. "the thing is that he needs a lot of loving.	Teď je vám 61 a je čas se usadit. "Jde o to, že potřebuje hodně lásky."
neutral uses	fact	vzhledem	We never expected John that er they would just change it like that in the light of the fact that they're going to change local government across the country anyway.	Nikdy jsme neočekávali, že John, že er, že to tímto způsobem změní vzhledem k tomu, že stejně změní místní vlády po celé zemi.
neutral uses	phenomenon	fenoménu	"[…].Later, mummy herself becomes the prototype for her aspirations and a little envy begins to creeps in." It is this supposed envy of mummy that is the most dangerous aspect of the new phenomenon.	„[…]. Později se maminka sama stala prototypem jejích tužeb a malá zá vist se začala vkrádat ." Právě tato domnělá závist k mamince je nejnebezpečnějším aspektem nového fenoménu.
neutral uses	case			
neutral uses	business			
neutral uses	point			
causal uses	reason	důvod	But Stehli has a good reason to write her story.	Ale Stehli má dobrý důvod napsat svůj příběh.
causal uses	cause			
causal uses	ground			
causal uses	result	výsledkem	When women are young, their careers and pension rights would be damaged if they took time off for their children, so grandmothers must help. The result is that you often know your grandchildren better than your own children!	Pokud jsou ženy mladé, jejich kariéra a nárok na důchod by byly ohroženy, kdyby si vzaly volno pro své děti, a proto musí pomoci babičky. Výsledkem je, že často znáte své vnoučata lépe než své vlastní děti!

Figure 7. Shell nouns in the Czech-English translations

3.1.4 Translation data from English to German

The translation of shell nouns in the German translations was based on the online version of the Langenscheidt Englisch-Deutsch Wörterbuch dictionary (https://de.langenscheidt.com/englisch-deutsch/), which not only provides word-to-word translation but also incorporates data from the German-English parallel corpus. The translation machine DeepL was also used as an aid to translate the target expressions from English to German. Part of the translation results, in which some language used seemed unnatural, were further revised.

The most frequent revision occurred with the English sentence structure "NP + is that…". This structure is rarely used in German; therefore, the translation process of the shell nouns involved a change of word class. For example, the shell noun "tragedy" is literally translated into "*Tragödie*" (see (7b)). However, it was more natural to shift the shell noun

to the adjective "*tragisch*" for this structure (see (7c)).

(7) (a) "Roy's Cornes's <u>tragedy is that</u> he was born with hemophilia."

(b) "*<u>Die Tragödie</u> von Roy Cornes <u>ist, dass</u> er mit Hämophilie geboren wurde*."

(c) "*<u>Es is tragisch, dass</u> Roy Cornes mit Hämophile geboren wurde*."

Some shell nouns do not exist in the German language and must be translated into an adjective or verb form. For example, "rush" can only be used as the verb "*beeilen*" in German, "guess" can only be used as the verb "*raten*," and "bet" can only be used as the verb "*wetten*," just to name a few. This linguistic phenomenon is especially interesting and should not be viewed as a counterexample of shell 'noun' because by definition 'noun', although is a commonly acceptable category, it is not realized in the same way in many languages. The fact that a noun in many languages is best translated as adjective or verb form in another language shows the best evidence for the existence of cross-language difference. This is also the goal of making a cross-language database. However, since we started from Schmid's terminology, we will not change the terminology we proposed for this work.

One difficulty in the translation process was reflecting the language's formality in the original text while preserving its word class. "*Nominalstil* (nominal structure)" is strongly connected to academic writing in German. In some cases, in which shell nouns involve the process of nominalization but are not used in a scientific scenario, German tends to use its verb (see (8b) and (8c)) or adjective/adverb (see (9b) and (9c)) form (both underlined).

(8) (a) "…his <u>refusal</u> to kowtow to the prevailing political currents did not make him a popular figure…"

(b) "…*seine <u>Ablehnung</u>, sich den vorherrschenden politischen Strömungen zu beugen,machte ihn nicht zu einer beliebten Figur*…"

(c) "…*dass er <u>ablehnte</u>, sich den vorherrschenden politische Strömungen zu beugen, machte ihn keinen beliebten Figur*…"

(9) (a) "…its <u>reluctance</u> to maintain investment and to diversify markets and products as rapidly as the dairy industry cannot be divorced from this relationship."

(b) "…*ihre <u>Zurückhaltung</u> bei der Aufrechterhaltung von Investitionen und bei der raschen Diversifizierung von Märkten und Produkten wie in der Milchwirtschaft kann von dieser Beziehung nicht losgelöst werden*."

(c) "…*<u>Ungern</u> hielt sie Investionen aufrecht und sie diversifizert <u>widerwillig</u> Märkte und Produkten wie in der Milchwirtscht, was von dieser Beziehung nicht losgelöst werden kann*."

3.1.5 Translation data from English to Vietnamese

To translate shell nouns from English to Vietnamese, we first used the *Cambridge* dictionary to get a detailed English explanation. After that, the shell nouns were translated into Vietnamese using the tratu website (http://tratu.soha.vn/)—an open online professional Vietnamese dictionary. Besides the word meaning, this website also provides bilingual collocations in different languages. Figure 8 illustrates the website's display after importing the shell nouns.
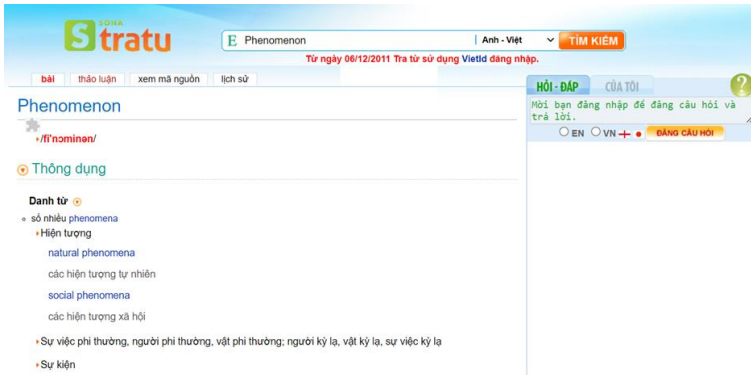
Figure 8. Translation of English shell nouns into Vietnamese

After ensuring the word's meaning, we inputted the data into Column C of the Excel file shown in Figure 9. In terms of sentence examples, instead of literal translation, we consulted some examples in the online Glosbe Dictionary (https://vi.glosbe.com/), and then translated and modified the target expressions to suit the context. For example, "phenomenon" (in red font), a polysemous word in English, can be translated as "*hiện tượng*" or "*sự kiện*." In this case, its meaning as "*hiện tượng*" (in red font) made the sentence clearer and more natural.

Figure 9. Shell nouns in the Vietnamese-English translations

3.2 Creation of the NCCU Shell Nouns Cross-Language Parallel Corpus (NCCU-SNPC) Interface

Step 3: Arrangement of cross-language data and establishment of the parallel corpus interface

(*https://nccu-eng.github.io/Shell-Nouns-Study/*)

To share the prepared the NCCU-SNPC openly, we leveraged the GitHub pages service to host all assets, including the collected materials and a jQuery-based user interface, to investigate data across languages interactively. This user interface provides a table view (see Figure 10) that shows the comparisons of shell nouns in different languages, an integrated search bar on the right-hand side for quick searches of specific words and terms, and a column-mask function to assist users in focusing on the preferred language.

Figure 10. The interface of the NCCU Shell Nouns Cross-Language Parallel

Corpus (NCCU-SNPC)

A flexible and easy-to-maintain framework was built on this website. All members of our research team can add new materials to the existing corpus, after compiling the data provided. A simple-to-use tool in the "*scripts*" folder is also provided in the GitHub software repository, and it generates the data source, located in the "*corpus*" folder, in a table view on the main page. After the system receives an updated corpus, the website will automatically deploy the latest version at [https://nccu-eng.github.io/Shell-Nouns-Study/].

4. Future Improvements to the NCCU Shell Nouns Cross-Language Parallel Corpus (NCCU-SNPC)

The NCCU-SNPC was built to explore the use of shell nouns between languages to explicitly understand the roles of shell nouns in context. However, there are still some improvements and adjustments that need to be made. First, the translated target words do not have clear divisions, making it difficult for users to discern the differences between similar translated words for selection. This can be improved by enhancing the display so that it includes categories that would make the target word

entered more clear. In this way, the data shown would be from the specific category that users chose to explore the outcome of the translations. Second, the interface of the system may not be easy to use for beginners. For example, the way the columns are displayed may not be user-friendly for those who are not familiar with the system. This can be improved by reformatting the columns into a list of selections to indicate which information should be displayed in the outcome section. Moreover, tutorial videos would be useful for users to understand how to apply this system to their needs and usage. Guidance on how to use the NCCU-SNPC is critical for users to understand how to operate each function and find the data they need.

All the collected data in the NCCU-SNPC is from Schmid's research examples. However, this data may be outdated in modern society. Therefore, it is important to collect a variety of real-life situational contexts to explore how shell nouns are used in real-life situations in modern society and in different cultures. Moreover, the collected data only focuses on translating the English context into other languages. Even though intertranslation between languages is critical, the source context is also a factor that influences translation strategies, by either contexts or translators' language proficiency and cultural backgrounds. Still, the collected data in the NCCU-SNPC is important in understanding how shell nouns are translated and interpreted between languages. To display the shell nouns more clearly, it is essential to show the translated shell nouns with different markings, such as using colored or bold font or underlining them, to show where shell nouns are presented in the translated the target expressions. Moreover, there are still some contexts waiting to be translated that are also important for analysts to explore how shell nouns are presented and translated between languages. As for the data output, download functions using Excel, Word, or txt files would be helpful for linguists to make further analyses, such as by marking their meanings for further discussion in research.

5. Conclusion and the Applications of the NCCU Shell Nouns Cross-Language Parallel Corpus (NCCU-SNPC)

In recent years, various researchers have developed corpora that detect and categorize terms to understand the meanings of shell nouns in context. These terms can be used as suggestions and sources for AI learning and

machine training. However, corpora for the translation of shell nouns are few to none. Therefore, the NCCU Shell Nouns Cross-Language Parallel Corpus (NCCU-SNPC) was designed to collect the English target expressions from Schmid (2000) and translate these target expressions into other languages to explore how shell nouns are interpreted between languages. The development of the NCCU-SNPC should shed light on directions for future studies exploring shell nouns' important role in connecting discourse to explain meanings in either English or other languages. Our target expressions could contribute to investigating shell nouns in various languages. It is unlike any translation tools or Google translator because our concern was to know how shell nouns could be translated into different languages.

In the research context, researchers can apply the NCCU-SNPC to find out discourse relations before and after the shell nouns to understand what roles shell nouns play in the context to explain what meanings are delivered by the speakers clearly. Since the collected data includes many languages, tagged data can be used to understand further which words are often used as shell nouns in one language or between languages. The categorized data can be used in pedagogy or translation to suggest which lexical choices are usually applied in the target context or between languages and their functions in discourse relations and grammatical patterns.

In pedagogical applications in the language learning curriculum, language teachers, for example, can use the NCCU-SNPC for classroom activities to introduce the meanings of shell nouns. After explaining the shell nouns, teachers can use the database to guide language learners in exploring which words or target expressions frequently appear as shell nouns. These target shell nouns can be helpful for language learners in understanding the meaning of the context. In other words, language learners can acquire the ability to identify shell nouns in reading contexts or use shell nouns as cohesive device to link ideas in discourse. These target shell nouns can also be used as a category for learning the meanings behind the shell nouns. Moreover, language learners can find out how shell nouns are translated between languages to see how the delivered information is translated through discourse relations. These methods can help language learners to explore their language learning process and language exchange between languages and their mother tongue.

If shell nouns contain meanings to be drawn from the context, they require high discourse-level context. We hope to see if the same group of shell nouns can be translated into other minority languages such as the Austronesian languages.

Bibliography

Ahn, Mikyung, and Foong Ha Yap. "'That Being so, but…': An Analysis of Korean Kunyang as a Marker of Speaker's Attenuated Divergent Stance." *Journal of Pragmatics*, vol. 160, 2020, pp. 31-43.

Aktas, Rahime Nur, and Viviana Cortes. "Shell Nouns as Cohesive Devices in Published and ESL Student Writing." Journal of English for Academic Purposes, vol. 7, no.1, 2008, pp. 3-14.

Benítez Castro, M. Á. *Formal, syntactic, semantic and textual features of English shell nouns*. Universidad de Granada, 2014.

Biq, Yung-O. "People, Things and Stuff: General Nouns in Spoken Mandarin." *Concentric: Studies in Linguistics*, vol. 30 no. 1, 2004, pp 41-64.

Carretero, Marta. "Noun Phrases as Expressions of Evidentiality: An Analysis of Four English Abstract Nouns and their Spanish Equivalents." *Kalbotyra*, vol. 69, no. 69, 2017, pp. 29-.

Charles, Maggie. "Argument or evidence? Disciplinary Variation in the Use of the Noun That Pattern in Stance Construction." *English for Specific Purposes*, vol. 26, no. 2, 2007, pp. 203-18.

Cuenca, Maria-Josep. "The Use of Demonstratives and Context Activation in Catalan Parliamentary Debate." *Discourse Studies*, vol. 16, no. 6, 2014, pp. 729-52.

Curry, Niall, and Pascual Pérez-Paredes. "Stance Nouns in COVID-19 Related Blog Posts." *International Journal of Corpus Linguistics*, 2021, pp. 469-97.

Dong, Min, and Alex Chengyu Fang. "Shell Nouns as Grammatical Metaphor Revealing Disparate Construals: Investigating the Differences between British English and China English Based on a Comparable Corpus." *Corpus Linguistics and Linguistic Theory*, vol. 17, no. 3, 2021, pp. 743–79.

Dong, Min, et al. "Shell Nouns as Grammatical Metaphor in Knowledge Construal: Variation across Science and Engineering Discourse." *Lingua*, vol. 248, 2020, pp. 102946-.

Fang, Alex Chengyu, and Min Dong. "Shell Nouns as Register-Specific Discourse Devices." *International Journal of Corpus Linguistics*, vol. 26, no. 2, 2021, pp. 219-47.

Flowerdew, John. "Signalling Nouns in Discourse." *English for specific purposes*, vol. 22, no. 4, 2003, pp. 329-46.

---. "Use of Signalling Nouns in a Learner Corpus." *International Journal of Corpus Linguistics*, vol.11, no. 3, 2006, pp. 345-62.

---. Use of signalling nouns across L1 and L2 writer corpora. *International*

Journal of Corpus Linguistics, 15.1 (2010): 36-55.

Flowerdew, John, and Richard W. Forest. *Signalling nouns in academic English*. Cambridge University Press, 2015

---. *Signalling Nouns in English: A Corpus-based Discourse Approach*. Cambridge: Cambridge University Press, 2017

Francis, Gill. *Anaphoric Nouns*. Birmingham. University of Birmingham Printing Section, 1986

Fujii, Seiko. "Incipient Decategorization of MONO and Grammaticalization of Speaker Attitude in Japanese Discourse." *Pragmatics and Beyond New Series*, 2000, pp. 85-118.

Halliday, Michael Alexander Kirkwood, and Ruqaiya Hasan. *Cohesion in English*. Longman, 1976

Hsieh, Chen-Yu Chester. *Interactional functions of Chinese shell-noun expressions: A study on wentishi, shishishang, zheyang(zi) and shemeysi.* 2012. Graduate Institute of Linguistics, National Taiwan University, Master's Thesis

---. "From Turn-Taking to Stance-Taking: *Wenti-shi* '(the) Thing is' as a Projector Construction and an Epistemic Marker in Mandarin Conversation." *Journal of Pragmatics*, vol. 127, 2018, pp. 107-24.

---. "Meaning in Repair: The Abstract Noun *Yisi* 'Meaning/Intention' in the Management of Intersubjectivity in Mandarin Conversation." *Taiwan Journal of Linguistics*, vol. 18, no. ,2020, pp. 39-88.

Hundt, Marianne, and Rahel Oppliger. "(The) Fact is…/*(Die) Tatsache ist*… Focaliser Constructions in English and German are Similar but Subject to Different Constraints." *International Journal of Corpus Linguistics*, vol. 27, no. 1, 2022, pp. 1-30.

Kim, Shin-Sook. "Noun complements and clause types in Korean (and Japanese)." *Japanese/Korean Linguistics*, vol. 18, 2011, pp. 278-90.

Kolhatkar, Varada, and Graeme Hirst. *"Resolving Shell Nouns." Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 499-510.

Kou, Xin. "The Discourse Functions of Shell Nouns in Mandarin: A Genre-Based Study in Popular and Professional Science Articles. *Workshop on Chinese Lexical Semantics*, 2022, pp. 230-40.

Lakoff, George, and Mark Johnson. "The Metaphorical Structure of the Human Conceptual System." *Cognitive science*, vol. 4, no. 2, 1980, pp. 195-208.

Lin, Melissa Shih-hui. *Pragmatické Užití Instrumentálu v Češtině: Z Hlediska L2 Pedagogiky = Pragmatic use of the instrumental case*

    in Czech: from the perspective of L2 Pedagogy. Taipei: Royal Orchid International, 2023.

Lin, Melissa Shih-hui, and Karel Šebesta. "Czech instrumentals as shell nouns in L2 pedagogy." *Journal of Slavic Languages*, vol. 28, no. 1, 2023, pp. 61-81.

Liu, Qingrong, and Liming Deng. "A Genre-Based Study of Shell-Noun Use in the N-Be-That Construction in Popular and Professional Science Articles." *English for Specific Purposes*, vol. 48, 2017, pp. 32-43.

Mazzi, Davide. "Grounds and Reasons: Argumentative Signals in Judicial Texts." *Linguistica e Filologia*, vol. 20, 2005, pp. 157-78.

---. "'This Argument Fails for Two Reasons…': A Linguistic Analysis of Judicial Evaluation Strategies in US Supreme Court Judgments." *International Journal for the Semiotics of Law—Revue internationale de Sémiotique juridique,*, vol. 23, no. 4, 2010, pp. 373-85.

Rosen, Alexandr, et al. *The InterCorp Corpus—Czech*, version 11 of 19 October 2018. 2018, http://www.korpus.cz.

Roze, Charlotte, et al. "Identification of Shell Nouns, Signals of Discourse Organisation (*Identification des Noms Sous-spécifiés, Signaux de l'Organisation Discursive*)[in French]." *Proceedings of TALN 2014* vol. 1, Long Papers, 2014, pp. 377-88.

Schmid, Hans-Jörg. "Cognitive Effects of Shell Nouns." *Amsterdam Studies in the Theory and History of Linguistic Science Series*, vol. 4, 1999, pp. 111-32.

---. *English Abstract Nouns as Conceptual Shells.* Mouton de Gruyter, 2000.

Swales, John M. "Metatalk in American Academic Talk: The Cases of Point and Thing." *Journal of English linguistics*, vol. 29, no. 1, 2001, pp. 34-54.

Tahara, Nobuko. *The use of shell nouns in Japanese and American student writing.* 2017. University of Birmingham, Doctoral Dissertation.

Takara, Nobutaka. "The Weight of Head Nouns in Noun-Modifying Constructions in Conversational Japanese." *Studies in Language*, vol. 36, no. 1, 2012, pp. 33-72.

Vergaro, Carla. *Illocutionary Shell Nouns in English*. Peter Lang AG, 2018

Vergaro, Carla, and Hans-Jörg Schmid. "Do the Meanings of Abstract Nouns Correlate with the Meanings of Their Complementation Patterns?: A Case Study on English Commissive Shell Nouns." *Pragmatics & Cognition*, vol. 24, no. 1, 2017, pp. 91-118.

Yang, Wei-Ting. *Shell Nouns in Mandarin Persuasive Discourse: A Corpus-based Move Analysis and the Pedagogical Implications*. 2022.

Address for correspondence

Siaw-Fong Chung
Department of English
National Chengchi University
No.64, Sec. 2, ZhiNan Rd.
Wenshan Dist.
11605 Taipei City
Taiwan

siawfongchung@gmail.com