ADRIAN JAN ZASINA
Charles University, Institute of Czech and Deaf Studies
adrian.zasina@ff.cuni.cz
ORCID: 0000-0001-9348-5833

The Representation of Taiwan in Czech Online Media Discourse between 2017 and 2020

Abstract

This paper examines the representation of Taiwan in Czech online media discourse between 2017 and 2020 using a corpus-based approach to discourse analysis. Drawing on the ONLINE1 corpus, which includes news, social media, and discussion data, the study analyses the collocational profiles of the noun *Tchaj-wan* ("Taiwan") and the adjective *tchajwanský* ("Taiwanese"). The findings show that *Tchaj-wan* is predominantly associated with political and geopolitical contexts, referring to international relations, official visits, and East Asian countries, while *tchajwanský* collocates mainly with business and technology terms, reflecting Taiwan's role in economic cooperation. Across media types, Taiwan is represented as both a democratic political partner and a technological ally. The study highlights a predominantly positive image of Taiwan in Czech online media discourse, which reflects the strengthening of Czech–Taiwanese relations during the period under examination.

Keywords: collocation analysis, corpus linguistics, Czech–Taiwanese relations, discourse analysis, media discourse, Taiwan

Introduction

Taiwan is a state recognised mainly by island countries and by one European state, the Holy See, where the Republic of China (ROC) maintains an embassy as of September 2025 https://www.taiwanembassy. org/). Although the ROC was recognised as a state after the Second World War and was even a member of the United Nations until 1971, its seat was then taken by the People's Republic of China (PRC) (Tanaka 2022: 125). Since that time, official recognition of Taiwan has gradually declined. However, the events surrounding COVID-19, originating in China, changed the situation and encouraged many countries to

deepen their relations with Taiwan. China's actions, such as its mistreatment of Hong Kong and Xinjiang (Chen: 2023: 38), have prompted many states to shift their focus from China towards Taiwan.

Between 2017 and 2020, the Czech Republic also renewed and strengthened its interest in Taiwan. For instance, the Mayor of Prague, Zdeněk Hřib, officially visited Taipei in 2019, and the President of the Czech Senate, Jaroslav Kubera, had planned to visit Taiwan before his death. Relations were further reinforced during the COVID-19 crisis, when Taiwan delivered medical supplies free of charge to the Czech Republic. This was followed by the official visit of the new President of the Czech Senate, Miloš Vystrčil, in September 2020. These events resonated with the general public and were widely discussed in the media.

This paper therefore seeks to examine the image of Taiwan in Czech online media between 2017 and 2020 and to comment on their typical thematic focus. For this purpose, it employs collocation analysis grounded in linguistic discourse analysis. This study addresses the following research questions:

- What are the most frequent collocates of the lexemes *Tchaj-wan* ("Taiwan") and *tchajwanský* ("Taiwanese") in Czech online media discourse between 2017 and 2020?
- How do these collocational patterns differ across the three data sources (discussions and fora, social media, and news)?
- What do these differences reveal about the representation of Taiwan and the nature of Czech–Taiwanese relations during this period?

The following section briefly introduces linguistic discourse analysis and outlines several examples of studies on discourse, media, and politics. After presenting the data, the research methodology is described. The results of the collocation analysis are then presented, including a comparison of the three different data sources. This is followed by a discussion of the main findings and concluding remarks.

Discourse analysis, media, and politics

In general, discourse analysis is an interdisciplinary perspective that enables researchers to uncover how social reality is constructed and maintained through language, social behaviour, and social perspectives (Sedláková 2015: 426–428). Combining discourse analysis with a corpus-based approach is not a simple quantitative procedure; rather, it involves numerous researcher choices and clearly defined analytical methods (Baker 2023: 221). Over time, this combination has crystallised into a standalone subdiscipline known as corpus-assisted discourse studies (CADS), which enables the analysis of an entire corpus rather than a single text that may reflect only a small and potentially biased part of the discourse (Gillings *et al.* 2023: 1). Analytical procedures may vary, and no universal guide exists for conducting discourse analysis (Sedláková 2015: 438).

In the Czech context, discourse analysis of media has been carried out, among others, by Renáta Sedláková (2015), who applies discourse analysis in the social sciences. With regard to corpus-based approaches to media, one of the first studies analyses media discourse through corpus-based representations of "Others" in relation to national minorities and immigrants in the Czech Republic (Fidler 2016). Another study examines adjectival collocations relating to minority groups (Roma, Vietnamese, and Ukrainian people) in the Czech Republic, showing that Ukrainians are more frequently

represented in a positive light than the other two groups (Elmerot 2019)and a corpus analysis of adjacent words may reveal how individuals or groups of people other than the sender (writer or speaker.

More recent analyses provide insights into other social groups. For instance, Anna Strakhova (2024) investigates Facebook discussions about three synonyms for *black person*, confirming that neutral and negative adjectives predominate over positive ones. The analysis of comments points to numerous xenophobic discussions, often accompanied by aggressive language and dehumanization of the target group based on ethnicity. Similarly, Irene Elmerot (2025) analyses sentiments towards three groups: nationalities, occupations, and gender. From a diachronic perspective, the results show that white-collar men represent the norm, while all other groups are represented as deviations from it. Irene Elmerot (2025: 16–19) also relates country names and their inhabitants to Gross National Income (GNI), noting that positive attitudes towards particular groups increase with higher GNI levels.

The issue of Taiwan is not commonly addressed in linguistics, whereas it is a frequently discussed topic in political science. Most often, it concerns the legal status of Taiwan as an independent territorial entity (Lin *et al.* 2022) or its relations with China and the geopolitical dynamics in the Taiwan Strait (Tanaka 2022). However, there are also studies devoted specifically to Czech–Taiwanese relations (Ian Tsung-Yen Chen 2023; Wen-Yu Chen 2023) which are key to understanding the broader context of political events relevant to linguistic analysis.

The present study combines media and political science with linguistic discourse analysis to shed light on the image of Taiwan through collocational analysis. It does not examine the representation of a social group or an individual, but rather that of a country name and lexical items denoting its origin. The study focuses on different data sources, analysing their similarities and dissimilarities. Therefore, this analysis shares features with political discourse studies (see Galasiński, Skowronek 2001), which examine issues concerning a country and its politics. An analysis focusing on one specific country has not yet been conducted, and these issues are therefore the subject of this study.

Data

For this analysis, the ONLINE1 corpus (Cvrček, Procházka 2020) was used. ONLINE1 is a monitoring corpus designed to capture evolving Czech internet content, including journalism, discussions, fora, and social media. It covers data from June 2017 to March 2021 and, in addition to online journalism, includes material from social media (Instagram, Facebook, Twitter), online discussions, and fora. Texts for the corpus were supplied by Dataweps (https://www.dataweps.com/).

The reason for using the ONLINE1 corpus was to capture a period rich in developments relating to Czech–Taiwanese relations. Another motivation was the availability of social media data, which are included only in this release. Social media platforms (Facebook, Twitter, Instagram) function differently from journalistic texts in terms of information dissemination; the most frequently reshared posts usually indicate greater public interest in a topic. Because Instagram data are available only up to December 2020, the present study is limited to the years 2017–2020.

The corpus contains more than seven billion tokens, summarised in Table 1, which presents both absolute and relative frequencies expressed as numbers of tokens (a token being a string of characters separated by spaces) for each year and each media source. Within individual media types, absolute

frequencies vary considerably; therefore, relative frequency expressed in instances per million (i.p.m.) is also included to allow for comparisons across groups of differing absolute sizes.

Table 1. ONLINE1 corpus structure by years.

Year	Frequency	Media source						Total
		Discussions	Facebook	Fora	Instagram	News	Twitter	
2017	absolute	212,374,784	631,737,259	166,483,383	34,711,842	313,649,929	248,646,999	1,607,604,196
	i.p.m.	132,106.39	392,968.16	103,559.93	21,592.28	195,103.95	154,669.29	1,000,000
2018	absolute	237,738,511	277,594,211	169,801,930	123,094,426	470,612,538	277,229,494	1,556,071,110
	i.p.m.	152,781.26	178,394.30	109,122.22	79,105.91	302,436.40	178,159.91	1,000,000
2019	absolute	170,184,084	264,799,566	111,245,663	233,080,049	477,988,531	317,443,283	1,574,741,176
	i.p.m.	108,071.15	168,154.34	70,643.78	148,011.66	303,534.66	201,584.42	1,000,000
2020	absolute	270,528,478	635,927,253	96,860,871	50,246,753	460,970,713	463,678,300	1,978,212,368
	i.p.m.	136,754.01	321,465.61	48,963.84	25,400.08	233,023.88	234,392.58	1,000,000
2021[1]	absolute	60,485,450	506,887	20,028,504	0	109,177,764	145,885,060	336,083,665
	i.p.m.	179,971.41	1,508.22	59,593.80	0.00	324,852.93	434,073.64	1,000,000
TOTAL	absolute	951,311,307	1,810,565,176	564,420,351	441,133,070	1,832,399,475	1,452,883,136	**7,052,712,515**

This study examines two Czech lexemes: the noun *Tchaj-wan* "Taiwan" and the adjective *tchajwanský* "Taiwanese," along with all existing spelling variants of both lexemes (*Tchaj-wan, Taiwan, Tchajwan, Tchai-wan, Tajvan* and *tchaj-wanský, tchajwanský, taj-wanský, tajwanský, tajvanský, thaj-wanský, thajwanský, thajvanský, thai-wanský, thaiwanský, thaivanský, tai-vanský*), taking into account possible variation in upper- and lower-case spelling. Other names for Taiwan, such as *Republic of China (ROC)* or *Formosa*, are not examined due to their low frequency.

Searches were carried out using the KonText interface (Machálek 2020) at https://www.korpus.cz. A total of 105,830 hits for *Tchaj-wan* and 17,284 for *tchajwanský* were obtained[2].

Methodology

In the first step, the top 100 collocation candidates for each analysed lexeme were identified. The KonText tool was used to extract collocation candidates based on the LogDice[3] measure; candidates were searched for by lemma[4] rather than by word form. Collocates were identified within a span of two words to the left and two words to the right, ensuring the extraction of the most relevant collocates of the lexemes under

---

1   Data are provided till March 2021.

2   For comparison, the lexeme *Hongkong* "Hong Kong" appeared only 17,717 times.

3   LogDice is an example of an association measure used to identify collocations (http://wiki.korpus.cz/doku.php?id=en:manualy:kontext:kolokace&rev=1533545959). Accessed 22.09.2025).

4   A *lemma* is the canonical form representing all word forms of a given lexeme. For instance, *jet, jede, jel, jedou* are word forms of the lemma *jít* "to go."

examination. The minimum collocate frequency in the corpus, as well as the minimum frequency within the span, was set at three occurrences.

Afterwards, the collocates were divided into part-of-speech (POS) groups and categorised into semantic categories. For categorisation, only autosemantic words (nouns, adjectives, and verbs) were considered. Some collocation candidates received high LogDice scores due to their frequent repetition on social media, which may have biased the results. While some of these were connected with frequently discussed topics (*e.g. plicní* "lung," referring to lung ventilators donated by Taiwan during the COVID-19 outbreak), others were, for example, Twitter account names (*e.g. @ZdenekHrib*) repeatedly reposted with the same message. However, reposting is typical behaviour on social media; therefore, for clarity, these collocates are retained in Tables 1 and 2 and are marked in grey. Collocates assigned to more than one category are shown in italics.

Collocation candidates were assigned to seven categories based on their predominant context: business, cooperation, geolocation, politics, statehood, sport, and reporting. The business-related category includes business events, company names, or their products (*e.g. TSMC, polovodič* "semiconductor"). The cooperation category covers Czech–Taiwanese relations, mostly concerning Taiwanese assistance during the COVID-19 outbreak (*e.g. dar* "gift," *poděkovat* "to thank"). The geolocation category comprises geographical terms and neighbouring countries, mainly in Southeast and East Asia (*e.g. průliv* "strait," *Japonsko* "Japan"). The politics category includes political events, politicians' names, and actions (*e.g. delegace* "delegation," *návštěva* "visit"). The statehood category concerns the question of Taiwan's existence as a country (*e.g. samostatnost* "autonomy," *vlajka* "flag"). The sport category includes sporting events and athletes' names (*e.g. Šu-wej* "Su-wei," *spoluhráčka* "teammate"). The final category, reporting, covers reports on events concerning culture, travel, news, and related topics (*e.g. výlet* "trip," *oolong* "oolong tea").

Table 2 represents lemmas collocating with the lexeme *Tchaj-wan* across all seven semantic categories. The largest categories are politics and geolocation. These include the most prominent topics appearing during the years 2017–2020, such as visits by Czech representatives to Taiwan, the expulsion of a Taiwanese representative from a meeting at the Czech Ministry, President Tsai, tensions between China and Taiwan, Tibet, Hong Kong, the (non-)recognition of Taiwan as part of China, and Southeast and East Asian countries in the context of the economy and the COVID-19 situation.

Table 2. Semantic categories of lemmas collocating with the lexeme *Tchaj-wan.*

Semantic category	Collocates
business	**Nouns**
	delegace "delegation," Semiconductor, Manufacturing, *mise* "mission," součástka "component," *vztah* "relation."
	Verbs
	obchodovat "to trade," investovat „to invest."

Semantic category	Collocates
	Nouns
	dar "gift," Satanasi[5], poděkování "thanks," pomůcka "equipment," ventilátor "ventilator."
cooperation	**Adjectives**
	nezištný[6] "disinterested," bezplatný "free," slíbený "promised," ochranný "protective," plicní "lung," poskytnutý "provided".
	Verbs
	ne/poděkovat[7] "not/to thank," darovat "to donate."
	Nouns
	Tibet, Hongkong "Hong Kong," Korea, Singapur "Singapore," Čína "China," Macao, Japonsko "Japan," ČLR "PRC," Vietnam, Taiwan, ostrov "island," Tchaj-pej "Taipei," Tchaj-wan "Taiwan," Filipíny "Philippines," Peking "Beijing," Taipei, Thajsko "Thailand," Malajsie "Malaysia."
geolocation	**Adjectives**
	pevninský "mainland," nedílný "integral," jižní "south," čínský "Chinese," ostrovní "island."
	Nouns
	Vystrčil "Miloš Vstrčil," @Vystrcil, Kubera "Jaroslav Kubera," *delegace* "delegation," *návštěva* "visit," *cesta* "journey," činitel "official," prezidentka "(female) president," senát "senate," @SenatCZ, zástupce "representative," @ZdenekHrib, @PavelFischer, odmítnutí "rejection," *mise* "mission," @Telicka, *výlet* "trip," @KantorLumir, *vztah* "relation," @andpolescuk, vykázání "expulsion. "
politics	**Adjectives**
	Vystrčilův "Vystrčil's," doněcký "Donetsk," klukovský "boyish," demokratický "democratic. "
	Verbs
	navštívit "to visit," *letět* "to fly," *jet* "to go," přirovnávat "to compare," vystrčit[8] " to push out," obdržet "to receive," vypravit [se] "to go. "

---

5   A nickname appearing in the discussion section.

6   The collocation tool produced the lemma zištný, which occurred only in negated contexts. Therefore, the negated form is included in the table.

7   The collocate poděkovat appeared in the data both with and without negation. To indicate this, a slash with the negation prefix is used.

8   This collocate is mostly used as a noun referring to Miloš Vystrčil. However, a few examples show wordplay, where it is used as a verb but retains the meaning referring to Miloš Vystrčil.

Semantic category	Collocates
statehood	**Nouns**
	nezávislost "independence," samostatnost "autonomy," uznání "recognition," provincie "province."
	Adjectives
	samostatný "independent."
	Verbs
	ne/uznávat[9] "not/to recognise," ne/uznat "not/to recognise," považovat "to consider."
sport	**Nouns**
	Šu-wej "Hsieh Su-wei," Hsieh "Hsieh Su-wei."
reporting	**Nouns**
	cesta "journey," *návštěva* "visit," soutěž[10] "competition," *výlet* "trip," zemětřesení "earthquake," odlet "departure," WHO, impérium "empire."
	Adjectives
	antivirový "antiviral."
	Verbs
	navštívit "to visit," *letět* "to fly," syntetizovat "to synthesise," *jet* "to go," *odletět* "to fly away."

Table 3 presents lemmas collocating with the lexeme *tchajwanský* across all seven semantic categories. The largest categories are politics and business. In addition to the political events already mentioned in connection with *Tchaj-wan*, the politics category also includes the launch of a direct China Airlines flight between Prague and Taipei, the re-election of Taiwan's president, and military manoeuvres in the Taiwan Strait. In the business category, the emphasis is on Taiwanese companies, particularly those involved in IT technologies and the semiconductor industry, as well as their investments.

Table 3. Semantic categories of lemmas collocating with the lexeme *tchajwanský*.

Semantic category	Collocates
business	**Nouns**
	Foxconn, TSMC, Foxconn, Airlines, HTC, Asus, *Chino*[11], DigiTimes, CviLux, MediaTek, Bioorganic, přepravce "carrier," investor, výrobce "producer," Wistron, Computex, polovodič "semiconductor," SafeDX, Economic, investice "investment," Kymco, Chieftec, Sym, *mise* "mission," NCC, Giant, *zástupce* "representative," Acer.
	Verbs
	ozývat [se] "to contact."

---

9 The collocates *uznávat* and uznat appeared in the data both with and without negation. To indicate this, a slash with the negation prefix is used.

10 In the data, it appears as a hashtag without Czech diacritics.

11 In fact, it is part of the company name China Airlines. The word *China* was automatically lemmatised as *Chi*no.

Semantic category	Collocates
cooperation	**Nouns**
	Autoland, technolog,[12] ventilátor "ventilator," dar "gift," bohemista "student of Czech studies," převzetí "takeover," *protějšek* "counterpart."
	Verbs
	přednést "to present."
geolocation	**Nouns**
	Tchaj-pej "Taipei," průliv "strait," úžina "strait," metropole "metropolis," Taiwan[ský][13] "Taiwanese," Kao-siung "Kaohsiung," Taipei, Šanghaj "Shanghai."
politics	**Nouns**
	Cchaj "Tsai Ing-wen," prezidentka "(female) president," vykázání "expulsion," *aerolinka* 'airline,' Chung-I "Wang Chung-yi," Jou "Jou Si-kun," @EvaSelepova, *Chino*, *aerolinie* "airline," Tsai "Tsai Ing-wen," oblak "cloud," zásilka "shipment," diplomat, Vystrčil "Miloš Vystrčil," recepce "reception," Kennedy, delegace "delegation," špión "spy," Vystrčilův "Vystrčil's," zákonodárce "legislator," parlament "parliament," diplomacie "diplomacy," *mise* "mission," Joanne "Joanne Ou," čínština "Chinese," telefonát "phone call," velvyslanec "ambassador," špion "spy," zastoupení "representation," @KantorLumir, uvěznění "imprisonment," @SenatCZ, *zástupce* "representative," *protějšek* "counterpart," velvyslanectví "embassy."
	Adjectives
	pokrokový "progressive," znovuzvolený "re-elected," vzdušný "air."
	Verbs
	proplout "to pass through," ne/blahopřát[14] "not/to congratulate," přivítat "to welcome," prášit "to dust," vykázat "to expel."
statehood	**Nouns**
	vlajka "flag," separatismus "separatism."
	Adjectives
	tibetský "Tibetan."
sport	**Nouns**
	Sie "Hsieh Su-wei," spoluhráčka "(female) teammate," Čchan "Chan Yung-jan," parťačka "(female) partner," CPBL "Chinese Professional Baseball League," partnerka "(female) partner," Strýcová "Barbora Strýcová," tenistka "(female) tennis player."
reporting	**Nouns**
	aerolinka "airline," *aerolinie* "airline," podporovatelka "(female) supporter," oolong "oolong tea," skauting "scouting," Zantovsky "Michael Žantovský."

---

12 In fact, this refers to the company name Autoland Technology Co., Ltd.

13 This is an error in automatic lemmatisation, as the item functions as an adjective in context.

14 The collocate blahopřát appeared in the data both with and without negation. To indicate this, a slash with the negation prefix is used.

The noticeable difference between the collocates of the noun *Tchaj-wan* and the adjective *tchajwanský* lies in the business semantic category, where the collocates with the adjective focus more on Taiwanese companies and their investments, than on expressing a general business context as in the collocates of the noun. In terms of cooperation, besides references to support during the pandemic, cooperation with Taiwanese students of Czech studies also appears among the collocates of the adjective. The geolocation category emphasises regional situations and tensions between China and Taiwan among the noun collocates, whereas the adjective collocates place greater emphasis on Taiwanese geography in a more localised sense. Taiwanese statehood is discussed more in terms of independence among the noun collocates, while the adjective collocates underline negative connotations associated with separatism. For a more detailed examination, this categorisation is further applied in the analytical section, where it is explained in greater depth.

The ONLINE1 corpus includes distinctly different data sources, i.e. internet news, Facebook posts, Twitter posts, Instagram posts, web discussions, and fora. To better understand the influence of data source on collocation profiles, these were divided into three groups: (1) discussions and fora (hereinafter DIS), (2) social media (Facebook, Instagram, Twitter; hereinafter SOC), and (3) internet news (hereinafter NEWS).

The 50 most frequent collocation candidates were extracted for each group using the same procedure as for the top 100 collocates: selection based on the LogDice measure, the same minimum collocate frequency, and the same minimum span frequency of three occurrences. The calculation was again carried out by lemma, and collocates were identified within a span of two words to the left and two words to the right. Only autosemantic words (nouns, adjectives, and verbs) were taken into account.

As a result, 48 collocates were obtained for DIS, 48 for SOC, and 47 for NEWS for the lexeme *Tchaj-wan* after excluding other POS, and 50 collocates respectively for DIS, SOC, and NEWS for the lexeme *tchajwanský*. These collocates were then subjected to further analysis.

The following section discusses the comparison between the three groups, highlighting their similarities and differences in terms of the most frequent collocates. Each subsection also provides an in-depth analysis of the collocates in relation to the semantic categories, while commenting on events that may have influenced their occurrence.

Comparison of three different data sources

The comparison of the three groups, DIS, SOC, and NEWS, reveals both similarities and differences in terms of data sources. Tables 4 and 5 present the most frequent collocates for each group. Collocates with more than half of their occurrences resulting from reposting are marked in grey. However, these may differ from those highlighted in the previous section or across the three groups, as they are not always repeated in all of them. Collocates that are underlined represent lemmas appearing among the most frequent collocates in all three groups.

The lexeme *Tchaj-wan* has ten collocates that appear across all data sources (see Table 4): *Vystrčil, Tibet, Singapur, Hongkong, Vystrčilův, Korea, delegace, Kubera, Čína,* and *návštěva*. These fall into three semantic categories: geolocation, politics, and reporting. Half of them concern geolocation, referring to Southeast and East Asian countries and the so-called Asian Tigers, or highlighting economic and

technological development. Singapore and South Korea also appear in a positive context, particularly in relation to their successful handling of the COVID-19 pandemic. Taiwan, Tibet, and Hongkong are mostly discussed in terms of their international recognition as independent states. These countries frequently appear alongside China within the framework of the One China Policy, where commentary often addresses international economic relations and potential conflicts regarding declarations of independence.

Five collocates, *Vystrčil, Vystrčilův, delegace, Kubera,* and *návštěva*, belong to the semantic category of politics. The most frequent contexts relate to the visit of the President of the Czech Senate, Miloš Vystrčil, to Taiwan in September 2020. This visit was widely discussed in the media before, during, and after the event, in all data sources. A similarly prominent topic was Jaroslav Kubera, the former President of the Czech Senate, who had planned to visit Taiwan but passed away in January 2020 before the visit could take place. The lexeme *návštěva* also falls into the category of reporting. Beyond its political meaning, it appears in the context of travel to Taiwan, often in reports on Taiwanese customs and notable sightseeing opportunities.

The discussed topics can be considered independent of the data source, as they occur across all three groups and fall into three semantic categories: geolocation, politics, and reporting. Some lemmas included in the data were reposted multiple times, which led to their duplication and, consequently, to higher frequency and collocation scores. Nevertheless, these lemmas were retained in the analysis, given the nature of online data, which are often shared or re-shared by users when there is increased interest in a topic. To account for issues associated with such duplication, these lexemes are marked and commented on separately.

As might be expected, the greatest duplication occurs within the SOC group, where re-sharing behaviour is most prevalent (16 lemmas). However, the DIS and NEWS groups also contain duplicated lemmas (10 in DIS and 5 in NEWS). Since these duplicated lemmas are not shared across all three groups, they are discussed in detail within the respective subsections for each group.

Table 4. Comparison of data sources based on the most frequent lemmas collocating with the lexeme *Tchaj-wan.*

No	DIS			SOC			NEWS		
	lemma	*Freq*	*logDice*	*lemma*	*Freq*	*logDice*	*lemma*	*Freq*	*logDice*
1.	Satanasi	227	8.51	Vystrčil	2 562	9.26	Šu-wej	416	8.74
2.	syntetizovat	217	8.43	Vystrcil	963	8.39	Vystrčil	888	8.11
3.	antivirový	217	7.90	ne/poděkovat	1 581	7.36	Tibet	281	7.27
4.	činitel	487	7.19	nezištný	319	7.12	Macao	141	7.03
5.	Vystrčil	327	6.84	Tibet	375	7.07	nezávislost	553	6.93
6.	slíbený	174	6.76	delegace	473	7.05	Hongkong	295	6.82
7.	Tibet	136	6.54	Kubera	519	6.91	Tchaj-pej	111	6.75
8.	součástka	118	6.20	doněcký	282	6.90	Semiconductor	104	6.70
9.	ČLR	65	5.64	SenatCZ	258	6.75	Kubera	334	6.64
10.	zakročit	61	5.49	návštěva	2 106	6.58	delegace	255	6.62
11.	Singapur	70	5.44	prezidentka	329	6.38	Manufacturing	92	6.56
12.	Taiwan	57	5.34	klukovský	165	6.36	Singapur	166	6.40

No	DIS			SOC			NEWS		
	lemma	Freq	logDice	lemma	Freq	logDice	lemma	Freq	logDice
13.	Honkong	24	5.27	Vystrčilův	166	6.36	Hsieh	76	6.31
14.	uznávat	171	5.10	ZdenekHrib	296	6.36	Filipíny	112	6.31
15.	Hongkong	73	5.02	dar	724	6.30	Jüng-žan	66	6.15
16.	dar	238	4.92	cesta	5 531	6.09	čína	420	6.14
17.	Vystrčilův	21	4.92	bezplatný	299	6.06	Vystrčilův	69	6.04
18.	Hong	27	4.85	přirovnávat	175	6.05	pevninský	81	5.86
19.	Formosa	16	4.78	Hongkong	235	6.04	Vietnam	181	5.86
20.	Korea	163	4.75	Macao	129	5.98	Korea	370	5.85
21.	vystrčit	35	4.74	poděkování	288	5.93	Malajsie	80	5.83
22.	samostatnost	45	4.72	PavelFischer	175	5.91	Yung-jan	45	5.60
23.	delegace	58	4.70	darovat	432	5.81	uznávat	259	5.60
24.	Kuberová	17	4.64	letět	659	5.78	návštěva	1019	5.59
25.	Kubera	74	4.63	Singapur	154	5.72	Peking	175	5.56
26.	čína	139	4.63	ne/uznávat	328	5.70	Cooperative	43	5.54
27.	virus	221	4.62	senát	844	5.66	neoddělitelný	48	5.53
28.	luskoun	16	4.49	vystrčit	125	5.57	Thajsko	107	5.53
29.	obchodovat	61	4.44	Taiwan	126	5.53	Japonsko	264	5.49
30.	Vietnam	56	4.35	Korea	338	5.52	ČLR	73	5.46
31.	Čína	673	4.31	KantorLumir	85	5.48	Chuang	38	5.35
32.	provincie	56	4.30	soutez	98	5.47	ostrovní	57	5.28
33.	Tchaj-wan	53	4.26	odmítnutí	146	5.47	plicní	118	5.27
34.	zoranžovět	11	4.26	ochranný	467	5.44	univerziáda	37	5.26
35.	nedílný	26	4.25	ČLR	108	5.37	darovat	256	5.25
36.	Čankajšek	11	4.22	Čína	1 399	5.31	Taipei	35	5.18
37.	Filipíny	19	4.16	Telicka	135	5.28	navštívit	512	5.17
38.	tibet	12	4.15	andpolescuk	105	5.21	Čao-cchun	33	5.17
39.	pevninský	18	4.12	navštívit	536	5.14	ventilátor	122	5.15
40.	uznání	49	4.10	Taipei	68	5.14	Indonésie	56	5.00
41.	pomůcka	117	4.07	zástupce	804	5.13	zástupce	699	4.99
42.	návštěva	348	4.07	poskytnutý	99	5.12	cesta	2 497	4.96
43.	Japonsko	88	3.99	vypravit	121	5.08	Čína	1 044	4.93
44.	samostatný	114	3.99	impérium	103	5.06	tchajwanský	43	4.93
45.	ROC	9	3.98	představující	71	5.05	prezidentka	89	4.90
46.	čina	14	3.96	sinopsiscz	69	5.02	Tchaj-wan	90	4.85
47.	uznat	137	3.93	obdržet	254	4.98	samostatnost	59	4.84
48.	ostrov	144	3.89	ne/uznat	325	4.97			

Eleven collocates are shared across the three groups for the lexeme *tchajwanský* (see Table 5), i.e. *Foxconn, Autoland, Cchaj, úžina, TSCM, Tchaj-pej, Chino* (in fact *China*), *průliv, prezidentka,* and *tibetský*. They belong to four semantic categories: business, cooperation, geolocation, and politics. The business category includes Taiwanese company names such as Foxconn, Taiwan Semiconductor Manufacturing Company (TSMC), China Airlines, and China Times. These appear in Czech media in the context of the semiconductor industry, technological development, and corporate investment and turnover. China Airlines is mentioned in relation to the planned opening of a branch office in Prague. One company name, Autoland Technology, appears in the cooperation category. It is associated with assistance provided to the Czech Republic during the pandemic, as Autoland Technology delivered respirators free of charge as part of mutual cooperation.

Four collocates, *úžina, Tchaj-pej, průliv,* and *tibetský*, fall into the geolocation category. The lexeme *Tchaj-pej* most often appears as a reference to the capital of Taiwan in general, and more specifically in the context of the partnership agreement between the City of Prague and Taipei City. *Úžina* and *průliv* refer to the Taiwan Strait, which has been a source of tension between China and Taiwan since the end of the Chinese Civil War (Tanaka 2022: 122). This is largely due to Chinese manoeuvres in the strait as a show of force, which are countered by American manoeuvres in support of Taiwan's autonomy. The Taiwan Strait also appears in the context of meteorological conditions related to the typhoons that frequently pass through the area.

The lexeme *tibetský* mostly relates to the Tibetan and Taiwanese flags that were hung in windows opposite the hotel where the Chinese president stayed in Prague in 2016. This incident attracted media attention in connection with the unlawful police intervention and the removal of the flags. It was a manifestation of support for Tibetan and Taiwanese independence, demonstrating the solidarity of many Czech citizens with Taiwan and Tibet. Ultimately, the court ruled that the police intervention had been unjustified.

The last category, politics, includes the collocates *Cchaj, Chino* (in fact *China*), *prezidentka,* and *proplout*. The collocates *Cchaj* and *prezidentka* refer to former Taiwanese president Tsai Ing-wen, a representative of Taiwan's Democratic Progressive Party (DPP). She appeared in the Czech media in several contexts, the most prominent being the deterioration of relations between Beijing and Taipei following her election in 2016. She openly rejected rapprochement with mainland China. She also held talks with Czech representatives (Zdeněk Hřib, Miloš Vystrčil), which received extensive media coverage. Her congratulatory message to Donald Trump after his election was likewise reported, often in connection with a perceived weakening of relations with China. The collocate *Chino* is linked to China Airlines, which was used by Czech delegates travelling to Taiwan. The collocate *proplout* refers to manoeuvres in the Taiwan Strait, as discussed in the paragraph above.

The topics falling into the four semantic categories (business, cooperation, geolocation, and politics) can, as with the lexeme *Tchaj-wan*, be regarded as independent of the data source.

Lemmas exhibiting duplication occur more frequently with the lexeme *tchajwanský* than with *Tchaj-wan*. The SOC group contains most of these lemmas, with twice as many as in the *Tchaj-wan* group (31 lemmas). The NEWS and DIS groups contain 14 and 9 duplicated lemmas respectively. Within SOC, two lemmas are shared with the other groups but display a high number of duplicates only in SOC. Thus, lemmas showing duplication differ considerably across the groups.

Table 5. Comparison of data sources based on the most frequent lemmas collocating with the lexeme *tchajwanský*.

DIS			SOC			NEWS		
lemma	*Freq*	*logDice*	*lemma*	*Freq*	*logDice*	*lemma*	*Freq*	*logDice*
Foxconn	119	11.42	Autoland	284	10.08	Cchaj	283	9.96
Foxconn	28	6.98	Cchaj	132	8.89	Sie	274	9.47
Autoland	6	6.62	vykázání	186	8.70	Tchaj-pej	182	8.97
Cchaj	6	6.29	Foxconn	109	8.01	TSMC	175	8.79
úžina	9	5.77	prezidentka	571	7.95	Foxconn	180	8.71
skůtry	3	5.63	Airlines	163	7.69	průliv	253	8.45
grošík	3	5.51	technolog	154	7.59	spoluhráčka	96	8.03
reexport	4	5.47	Chung-I	40	7.34	Čchan	64	7.93
čínština	10	5.24	EvaSelepova	42	7.29	prezidentka	491	7.73
TSMC	7	5.21	CPBL	35	7.11	HTC	121	7.27
Vystrčilův	5	4.84	Tchaj-pej	43	6.91	úžina	55	7.20
ZF	3	4.78	průliv	83	6.85	Autoland	32	6.90
Tchaj-pej	4	4.68	Bioorganic	25	6.70	DigiTimes	30	6.83
virolog	11	4.55	ozývat	197	6.56	CviLux	28	6.82
Chino	7	4.27	znovuzvolený	27	6.56	parťačka	57	6.64
průliv	9	4.12	podporovatelka	28	6.50	proplout	41	6.57
prezidentka	32	3.93	aerolinka	124	6.47	MediaTek	27	6.46
proplout	3	3.75	TSMC	32	6.36	Asus	88	6.31
anabáze	3	3.67	oolong	20	6.33	Jou	33	6.25
Vyskoč	3	3.60	ventilátor	201	6.18	SafeDX	16	5.99
trpaslík	15	3.42	přepravce	39	5.91	Wistron	16	5.98
Airlines	6	3.35	aerolinie	30	5.89	partnerka	194	5.94
tibetský	8	3.20	pokrokový	68	5.88	Kao-siung	15	5.91
Asus	8	3.15	oblak	73	5.85	Chino	33	5.86
skřet	4	3.09	Jou	24	5.82	Economic	18	5.81
aerolinka	9	2.91	Zantovsky	14	5.76	Strýcová	58	5.78
Vystrčil	16	2.78	Tsai	13	5.68	Kymco	15	5.77
AZ	6	2.76	zásilka	169	5.65	Computex	15	5.76
kašpárek	4	2.72	skauting	28	5.64	bohemista	15	5.67
hongkongský	3	2.62	KantorLumir	12	5.51	aerolinka	70	5.64
čivava	4	2.56	ne/blahopřát	114	5.43	Chieftec	12	5.59
korejský	10	2.49	přivítat	173	5.42	NCC	12	5.58
KS	4	2.46	Sie	16	5.40	Chung-I	12	5.56
inaugurace	4	2.41	úžina	15	5.36	Joanne	13	5.47
delegace	8	2.22	Chino	23	5.35	Tsai	11	5.40
investice	57	2.12	dar	278	5.21	Sym	11	5.33

DIS			SOC			NEWS		
lemma	*Freq*	*logDice*	*lemma*	*Freq*	*logDice*	*lemma*	*Freq*	*logDice*
diplomat	12	2.03	Li-lin	9	5.20	aerolinie	20	5.29
Taiwan	3	1.76	Kennedy	31	5.14	vykázání	17	5.23
loterie	4	1.70	investor	262	5.10	polovodič	13	5.22
firma	236	1.59	uvěznění	16	5.09	recepce	53	5.21
protějšek	4	1.48	oolonzích	8	5.05	Giant	13	5.12
pozvání	8	1.40	přednést	33	5.05	Hingisová	10	5.12
přepis	3	1.23	tibetský	35	5.03	tenistka	65	5.03
investor	17	1.20	Lien	8	5.00	separatismus	11	5.02
sportovec	14	1.13	Taipei	9	4.96	metropole	125	5.02
provincie	5	1.10	prášit	22	4.94	Jüng-tchem	8	5.01
národnost	4	1.06	SenatCZ	22	4.86	PTT	8	4.99
vlajka	11	0.84	oolongy	7	4.85	tibetský	34	4.98
velvyslanec	7	0.81	wang	12	4.83	excellence	14	4.92
elektronika	4	0.69	proplout	12	4.82	36letý	10	4.90

This section has provided a general description of the lemmas included in Tables 4 and 5 and has commented primarily on the similarities between the individual data sources. The differences are discussed in the following three subsections, which focus on the typical collocates for the DIS, SOC, and NEWS groups.

Discussion and forum group

The DIS group contains fora and discussions beneath online press articles. It represents a dynamic mode of communication that occurs multidirectionally between participants (see Cvrček *et al.* 2021: 358–359). Consequently, this group typically reflects people's opinions on individual events. Most collocates appear in topics related to politics, geolocation, and business (see Table 6).

Table 6. Semantic categories of lemmas collocating with the lexemes *Tchaj-wan* and *tchajwanský* within the DIS group.

Semantic category	Collocates	
	Tchaj-wan	**tchajwanský**
	Nouns	
business	součástka "component," *delegace* "delegation."	Foxconn, Foxconn, skůtry[15] "scooters," reexport, TSMC, ZF, Airlines, Asus, AZ, investice "investment," firma "company," investor, elektronika "electronics."
	Verbs	
	obchodovat "to trade."	

---

15 The lemma is written in a non-standard spelling; the correct form should be skútr.

Semantic category	Collocates	
	Tchaj-wan	**tchajwanský**
cooperation	**Nouns**	
	dar "gift," Satanasi.	Autoland.
	Adjectives	
	slíbený "promised."	
geolocation	**Nouns**	
	Tibet (also tibet), Singapur "Singapore," ČLR "PRC," Taiwan, Hongkong (also Hong Kong), Formosa, Korea, Vietnam, Čína "China" (also čína, čina), Tchaj-wan "Taiwan," Filipíny "Philippines," Japonsko "Japan," ROC, ostrov "island."	úžina "strait," Tchaj-pej "Taipei," průliv "strait," tibetský "Tibetan," hongkongský "Hongkongese," korejský "Korean," Taiwan.
	Adjectives	
	pevninský "mainland," nedílný "integral."	
politics	**Nouns**	
	činitel "official," Vystrčil "Miloš Vystrčil," *delegace* "delegation," Kuberová "Věra Kuberová," Kubera "Jaroslav Kubera," Čankajšek "Chiang Kai-shek," návštěva "visit."	Cchaj "Tsai Ing-wen," grošík "penny," čínština "Chinese," Chino "China Airlines," prezidentka "(female) president," anabáze "anabasis," Vyskoč "Pepek Vyskoč," trpaslík "dwarf," skřet "goblin," aerolinka "airline," Vystrčil "Miloš Vystrčil," kašpárek "clown," čivava "chihuahua," KS "Comunist Party," inagurace "inauguration," delegace "delegation," diplomat, protějšek "counterpart," pozvání "invitation," velvyslanec "ambassador."
	Adjectives	
	Vystrčilův "Vystrčil's."	Vystrčilův "Vystrčil's."
	Verbs	
	vystrčit[16] "to push out."	proplout "to pass through."
statehood	**Nouns**	
	provincie "province," samostatnost "autonomy," uznání "recognition."	národnost "nationality," vlajka "flag," provincie "province"
	Adjectives	
	samostatný "independent."	
	Verbs	
	ne/uznat "to recognise."	
sport	**Nouns**	
		sportovec "sportsman"

---

16 This collocate is mostly used as a noun referring to Miloš Vystrčil. However, a few examples show wordplay, where it is used as a verb but retains the meaning referring to Miloš Vystrčil.

Semantic category	Collocates	
	Tchaj-wan	**tchajwanský**
reporting	**Nouns**	
	virus, luskoun "pangolin," pomůcka "equipment."	virolog "virologist," loterie "lottery," přepis "transcription."
	Adjectives	
	antivirový "antiviral."	
	Verbs	
	syntetizovat "to synthesise," zakročit "to intervene," zoranžovět "to turn into orange."	

This section focuses on collocates that are typical of the DIS group. In the business category, collocates for the lexeme *tchajwanský* occur more frequently than those for *Tchaj-wan*. Taiwanese investments in the Czech Republic are discussed in comparison with Chinese investments (collocates: *reexport, investice, firma, investor, součástka, obchodovat*), which, according to the commentaries,[17] surpass Chinese investment levels. The commentaries also refer to Taiwanese company names and their technologies (collocates: *Asus, elektronika, AZ, ZF, skútry*). A discussion concerning China Airlines (collocate: *Airlines*) notes that the company is of Taiwanese origin and appears in the context of Vystrčil's delegation to Taiwan (collocate: *delegace*).

All collocates within the cooperation category relate to a single event: Taiwan's assistance during the pandemic through the supply of ventilators (*dar, slíbený*). The company Autoland Technology produced the ventilators intended for the Czech Republic. The nickname *Santasi* appears as a duplicate in identical messages commenting on free Taiwanese assistance, in contrast to China, which did not provide anything free of charge.[18]

The geolocation category includes a list of Southeast and East Asian countries (collocates: *Vietnam, Filipíny, Japonsko*), referring either to events in the region or to countries (collocates: *Taiwan, Vietnam, Japonsko*) cited as examples of successful pandemic management. Other collocates (*ČLR, ROC, pevninský, ostrov, nedílný*) concern Taiwan–China relations, particularly Taiwan's status as a state and Taiwanese national identity. The country name *Taiwan* is also discussed in the context of Czech spelling, as it appears in two forms: *Tchaj-wan* and *Taiwan*. The former is typically used in official references to the Republic of China. The collocate *Formosa* appears both in its historical sense and as another designation for the island. The adjectives *hongkongský* and *korejský* indicate origin from those respective countries.

Within the semantic category of politics, many collocates refer to Vystrčil. His pejorative portrayal is evident in online discussions, where he is compared to various fictional characters (collocates: *Vyskoč, trpaslík, skřet, kašpárek, čivava*) or mocked for his supposed greed for Taiwanese money (collocate: *grošík*).

---

17 Corpus excerpt example: "(…) tchajwanské investice u nás do konce r. 2017 činily 5,7 miliardy Kč, zatím co ty čínské byly jen 410 miliónů Kč = 14 krát méně!" (Taiwanese investments in our country by the end of 2017 amounted to 5.7 billion CZK, while Chinese investments were only 410 million CZK = 14 times less!). https://www.korpus.cz/kontext/view?q=~investments. Accessed 19.09.2025.

18 Corpus excerpt example: "*Tchai-wan nám poslal "jen" 25 ventilátorů, ale Čína, které lezeme do zadku, zadarmo ani prd*" (Taiwan sent us "only" 25 ventilators, but China, up whose arse we are crawling, does not give a damn for free). https://www.korpus.cz/kontext/view?q=~Santas. Accessed 19.09.2025.

Further comments relate to his delegation to Taiwan and his pro-Taiwan stance. As an act of sympathy and support for democracy, he declared the sentence "I am Taiwanese" in Chinese during his official speech in Taiwan (collocates: *Vystrčil, Vystrčilův, Kuberová, delegace, čínština, anabáze*).

Another political issue concerns the expulsion of the Taiwanese ambassador from a meeting at the Czech Ministry of Industry and Trade, as well as discussions about the removal of a Taiwanese diplomat during the Prague mayor's New Year gathering (collocates: *diplomat, velvyslanec, inaugurace*). The next group of collocates (*činitel, protějšek, pozvání*) refers to international relations and tensions between China and the Czech Republic arising from the official state visit to Taiwan. There are also comments about the Communist Party (collocate: *KS*) operating in Taiwan, which seeks to unify China, and historical references to the state leader Chiang Kai-shek, who fled to Taiwan (collocate: *Čankajšek*). The collocate *proplout* appears in the context of an American flotilla passing through the Taiwan Strait.

Many discussions on Taiwan's statehood included collocates such as *provincie, samostatnost, uznání, národnost, vlajka, ne/uznat*. Contributors express their views on whether Taiwan should be recognised and what such recognition might imply for the future. The sport category is represented by a single collocate, *sportovec*, which also appears as a duplicate. It refers to comments on Taiwan's participation in the Olympic Games, where it competes as Chinese Taipei. The final category, reporting, mostly concerns the pandemic, with discussions on vaccination, viruses, medical equipment, and related topics (collocates: *virus, pomůcka, virolog, antivirový, syntetizovat, zakročit*). Other themes include animals donated to Prague Zoo (*luskoun*), Taiwanese cultural items such as the invoice lottery (*loterie*), and the transcription of Chinese names (*přepis*).

Social media group

The SOC group consists of data from social media platforms such as Facebook, Instagram, and the former Twitter. These data are dynamic, and the resharing of information is common. Consequently, this group contains the highest number of lemmas showing duplication. Most lemmas fall within the semantic categories of politics, reporting, and geolocation (see Table 7). There are no examples from the business or sport categories for the lexeme *Tchaj-wan*, and, conversely, no examples from the statehood category for the lexeme *tchajwanský*.

Table 7. Semantic categories of lemmas collocating with the lexemes *Tchaj-wan* and *tchajwanský* within the SOC group.

Semantic category	Collocates	
	Tchaj-wan	**tchajwanský**
	Nouns	
business		Foxconn, Airlines, Bioorganic, TSMC, přepravce "carrier," *Chino* "China Airlines," investor.
	Verbs	
		ozývat [se] "to contact."

Semantic category	Collocates	
	Tchaj-wan	**tchajwanský**
cooperation	**Nouns**	
	poděkování "thanks," @sinopsiscz, dar "gift."	Autoland, technolog, ventilátor "ventilator," dar "gift," Li-lin "Li-Lin Kuo."
	Adjectives	
	nezištný "disinterested," bezplatný "free," ochranný "protective," poskytnutý "provided."	
	Verbs	
	ne/poděkovat "not/to thank," darovat "to donate," *obdržet* "to obtain."	přednést "to present."
geolocation	**Nouns**	
	Tibet, Hongkong "Hong Kong," Macao, Singapur "Singapore," Taiwan, Korea, ČLR "PRC," Čína "China," Taipei.	Tchaj-pej "Taipei," průliv "strait," úžina "strait," tibetský "Tibetan," Taipei.
politics	**Nouns**	
	Vystrčil "Miloš Vystrčil," @Vystrcil, delegace "delegation," Kubera "Jaroslav Kubera," @SenatCZ, *návštěva* "visit," prezidentka "(female) president," @ZdenekHrib, *cesta* "journey," @PavelFischer, senát "senate," @KantorLumir, @Telicka, @andpolescuk, zástupce "representative."	Cchaj "Tsai Ing-wen," vykázání "expulsion," prezidentka "(female) president," Chung-I "Wang Chung-yi," @EvaSelepova, *areolinka* "airline," aerolinie "airline," pokorokový "progressive," oblak "cloud," Jou "Jou Si-kun," Tsai "Tsai Ing-wen," zásilka "shipment," @KantorLumir, ne/blahopřát "not/to congratulate," Kennedy "John Kennedy," uvěznění "imprisonment," Lien "Lien Chan," @SenatCZ, wang "Wang Mei-hua."
	Adjectives	
	doněcký "Donetsk," klukovský "boyish," Vystrčilův "Vystrčil's," odmítnutí "rejection."	znovuzvolený "re-elected."
	Verbs	
	přirovnávat "to compare," *letět* "to fly," vystrčit "to push out," *navštívit* "to visit," vypravit [se] "to go," *obdržet* "to obtain."	přivítat "to welcome," prášit "to dust," proplout "to pass through."
statehood	**Verbs**	
	ne/uznávat "not/to recognise," ne/uznat "not/to recognise."	
sport	**Nouns**	
		CPBL "Chinese Professional Baseball League," Sie "Hsieh Su-wei."
reporting	**Nouns**	
	návštěva "visit," *cesta* "journey," *letět* "to fly," soutěž[19] "competition," impérium "empire."	podporovatelka, *areolinka* "airline," oolong "oolong tea," oolonzích "oolong teas."
	Adjectives	
	představující "presenting."	*aerolinie* "airline," Zantovsky, sakuting, *Chino* "China Airlines," oolongy "oolong teas."
	Verbs	
	navštívit "to visit."	

19 In the data, it appears as a hashtag without Czech diacritics.

The business semantic category is primarily associated with the lexeme *tchajwanský*, where company names appear in relation to their products and investments (collocates: *Foxconn, Airlines, Bioorganic, TSMC*). There are also comments referring to China Airlines and Vystrčil's delegation to Taiwan, but China Airlines likewise appears in discussions about direct flights between Prague and Taipei. The collocates *investor* and *ozývat* show a high number of duplicates; they relate to Taiwanese investment in the Czech Republic and to a single reposted tweet containing information about a business activity.[20]

As in the previous subsection, all collocates within the cooperation category relate to Taiwan's support through the delivery of ventilators during the pandemic. The popularity of Z. Hřib's tweet about the Taiwanese donation[21] is particularly notable. Other reposted social media content concerns the businessman Li-Lin Kuo, who donated respirators. The geolocation category, in addition to topics similar to those in the DIS group, includes information about the pandemic situation in Asian countries (collocates: *Macao, Singapur, Korea*) and expressions of praise for Taiwan's management of the coronavirus crisis. Taiwan also appears in discussions about its legal status as a country, while *Tchaj-pej* occurs in references to the capital, the Taipei 101 tower, the zoo, or simply as part of an address.

The topics within the politics category overlap with those discussed above. However, no pejorative expressions about Vystrčil appear here. Based on the reposted collocates, it seems that the most prominent topics on social media concerned the official Czech visit to Taiwan and the expulsion of a Taiwanese representative from a meeting at the Czech Ministry of Industry and Trade. Attention was also given to the re-election of President Tsai. The question of Taiwan's statehood was associated with only two verbal collocates (*ne/uznávat* "to recognise," *ne/uznat* "to recognise"), which appeared both with and without negation, but only in combination with the noun *Tchaj-wan*. This indicates that public opinion is divided regarding the recognition of Taiwan.

The lexeme *tchajwanský* relates to two sporting events. The first concerns a baseball team in the Chinese Professional Baseball League (CPBL) and its results. The second relates to the tennis match between Barbora Strýcová and Sie Šu-wej. The reporting category includes information about tourism (collocates: *návštěva, cesta, letět, navštívit, aerolinka*), where tips and comments on travelling to Taiwan can be found. Taiwanese culture (collocates: *oolong, oolonzích, oolongy*) is represented by Taiwan's national tea, described as a high-quality product. The newspaper *Deník* organised a contest to win a trip to Taiwan, which contributed to the frequency of the collocate *soutěž*, appearing as a hashtag.

News

The NEWS group covers internet news sources, including, for instance, mainstream, tabloid, analytical–investigative, and anti-system media. Compared with the SOC group, it contains fewer collocates showing duplication and is less expressive than those in the DIS group found in discussions and forum comments.

---

20 @jirkkus: "Už včera se mi začali ozývat první tchajwanští investoři, kteří mají zájem o české nanotechnologie" (Yesterday, the first Taiwanese investors began contacting me; they are interested in Czech nanotechnologies); https://x.com/jirkakus/status/1302287296973148164/. Accessed 20.09.2025.

21 @zdenekhrib: "Plicní ventilátory z tchaj-wanského daru budou zachraňovat lidské životy v nemocnicích v Praze!" (Lung ventilators from a Taiwanese donation will save lives in hospitals in Prague); https://x.com/ZdenekHrib/status/1247877661604679681/. Accessed 20.09.2025.

In comparison with both of the other groups, it includes more collocates related to sport, while business, geolocation, and politics are also important (see Table 8).

Table 8. Semantic categories of lemmas collocating with the lexemes *Tchaj-wan* and *tchajwanský* within NEWS group.

Semantic category	Collocates	
	Tchaj-wan	**tchajwanský**
business	**Nouns**	
	Semiconductor, *delegace* "delegation," Manufacturing, *Cooperative.*	TSMC, Foxconn, HTC, DigiTimes, CviLux, MediaTek, Asus, SafeDX, Wistron, *Chino* "China Airlines," Economic, Kymco, Computex, Chieftec, NCC, SYM, polovodič "semiconductor," Giant, PTT, excellence "excellence."
cooperation	**Nouns**	
	ventilátor "ventilator."	Autoland.
	Adjectives	
	plicní "lung."	
	Verbs	
	darovat "to donate."	
geolocation	**Nouns**	
	Tibet, Macao, Hongkong "Hong Kong," Tchaj-pej "Taipei," Singapur "Singapore," Filipíny "Philippines," Čína (also čína) "China," Vietnam, Korea, Malajsie "Malaysia," Peking "Beijing," Thajsko "Thailand," Japonsko "Japan," ČLR "PRC," ostrovní "island," Taipei, Indonésie "Indonesia," tchajwanský "Taiwanese," Tchaj-wan "Taiwan."	Tchaj-pej "Taipei," průliv "strait," úžina "strait," Kao-siung "Kaohsiung," metropole "metropolis," tibetský "Tibetan."
	Adjectives	
	pevninský "mainland."	
politics	**Nouns**	
	Vystrčil "Miloš Vystrčil," Kubera "Jaroslav Kubera," *delegace* "delegation," *návštěva* 'visit," *Cooperative*, zástupce "representative," *cesta* "journey," prezidentka "(female) president."	Cchaj "Tsai Ing-wen," prezidentka "(female) president," Jou "Jou Si-kun," *Chino* "China Airlines," *areolinka* "airline," Chung-I "Wang Chung-yi," Joanne "Joanne Ou," Tsai "Tsai Ing-wen," *aerolinie* "airline," vykázání "expulsion," recepce "reception," Jüng-tchem "Lee Yung-te."
	Adjectives	
	Vystrčilův "Vystrčiľ's."	
	Verbs	
	navštívit "to visit."	proplout "to pass through."

Semantic category	Collocates	
	Tchaj-wan	**tchajwanský**
statehood	**Nouns**	
	nezávislost "independence," samostatnost "autonomy."	separtismus "separatism."
	Adjectives	
	neoddělitelný "inseparable."	
	Verbs	
	ne/uznávat "not/to recognise."	
sport	**Nouns**	
	Šu-wej "Hsieh Su-wei," Hsieh "Hsieh Su-wei," Jüng-žan (also Yung-jan) "Latisha Chan," Chuang[22] "Yu-Ting Huang," univerziáda "Universiade," Čao-cchun "Cheng Chao-tsun."	Sie "Hsieh Su-wei," spoluhráčka '(female) teammate," Čchan "Chan Yung-jan," parťačka "(female) partner," partnerka "(female) partner," Strýcová "Barbora Strýcová," Hingisová "Martina Hingis," tenistka "(female) tennis player."
reporting	**Nouns**	
	návštěva "visit," Chuang "Chia-Chun Chuang," *cesta* "journey."	bohemista "student of Czech studies," *areolinka* "airline," *aerolinie* 'airline."
	Adjectives	
		36letý "36-year-old."
	Verbs	
	navštívit "to visit."	

Business-related topics concern high-tech companies involved in the semiconductor industry (collocates: *Semiconductor, Manufacturing, TSMC, HTC, MediaTek, polovodič*), electronic components (*Foxconn*), automobiles (*Kymco, SYM*[23]), computer hardware (*Asus, Wistron*), and other companies (*Cooperative,*[24] *Cvilux, Chino,*[25] *Chieftec, Giant (Bicycles), NCC*[26]) and portals (*DigiTimes, Economic (Daily), PTT*) of Taiwanese origin. The media most often report on companies' new products, sales growth, or innovations in technology and production. News items also cover the Taiwanese Computex trade fair (collocate: *Computex*). The duplicates of the collocate *exelence* relate to the Taiwanese Excellence Gold Awards.

The cooperation semantic category, as in the two previous groups, concerns Taiwan's support in supplying ventilators to the Czech Republic. In terms of the geolocation category, besides a list of Southeast Asian countries and references to COVID-19, some collocates appear in the context of China's territorial claims in various parts of the South China Sea (collocates: *Filipíny, Vietnam, Malajsie, Indonésie*).

---

22 This collocate appears only once and was included in the data due to the identical spelling of the sportswoman's name Chia-Chun Chuang.

23 Sanyang Motor.

24 Taiwan Cooperative Bank (TCB).

25 China Airlines.

26 National Communications Commission (NCC).

The topic of tension between China and Taiwan is also present, with China regarding Taiwan as an integral part of its territory and accusing it of being a rebel province (collocates: *Čína, Peking, ostrovní*).

The official visits of Czech delegates to Taiwan and the expulsion of the Taiwanese ambassador from a meeting are the most prominent topics within the politics semantic category. In addition to the previous groups, a meeting between Miloš Vystrčil and representatives of the Taiwan Cooperative Bank is mentioned, concerning the establishment of a Taiwanese branch in the Czech Republic (collocate: *Cooperative*). The statehood category includes collocates relating to Taiwan's independence (*nezávislost, samostatnost*) and questions regarding its sovereignty (*neoddělitelný, ne/uznávat*). The duplicates of the collocate *separatismus* reflect Taiwanese tendencies framed as a threat of separatism.

In contrast to the two previous groups, the NEWS group comments much more extensively on sport. The focus is mainly on tennis matches and their players (*e.g. Hsieh, Yung-jan, Chuang, Strýcová, tenistka*). There is also information about the World Universiade held in Taiwan in 2017 (collocate: *univerziáda*). In addition to information about tourist visits to Taiwan or the Taiwanese China Airlines, reports include visits by Miloš Vystrčil or Václav Havel to Taiwanese students of Czech at National Chengchi University (collocate: *bohemista*). Furthermore, an incident involving a Taiwanese traveller who lost her life in the mountains appears as a duplicate (collocate: *36letý*).

Discussion and concluding remarks

This study has sought to examine how Taiwan is represented in the Czech online media captured in the ONLINE1 corpus during the period 2017–2020, and to identify the most common topics associated with frequent collocations. The results indicate that the prevailing themes in this period were: visits of Czech delegates to Taiwan; Taiwanese assistance during the COVID-19 crisis; and questions concerning relations between China and Taiwan, particularly in relation to Taiwan's statehood. The comparison between the three data groups – discussions and fora, social media, and news – reveals topics that can be considered either data-source-independent or data-source-dependent. Data-source-independent topics include:

- Southeast and East Asian countries and their handling of the COVID-19 pandemic,
- international recognition of Taiwan, Tibet, and Hong Kong as independent states,
- discussions concerning the One China Policy,
- the visit of the President of the Czech Senate to Taiwan,
- travel-related and cultural information about Taiwan,
- Taiwanese companies in the context of the high-tech industry and technological development,
- Taiwan's support provided to the Czech Republic during the pandemic,
- the Taiwan Strait as a strategic location,
- the display of Tibetan and Taiwanese flags opposite the Chinese president's hotel in Prague,
- the visibility of former Taiwanese president Tsai Ing-wen,
- China Airlines' business presence and the introduction of a new flight connection with Prague.

A comparison between the noun and adjective under examination also indicates a stronger emphasis on business within the collocates of the adjective, and a stronger emphasis on geolocation within the collocates of the noun. Both the noun and the adjective emphasises political issues.

The differences between the three groups show that the discussion-and-fora group is characterised by more opinion-driven expressions of people's views (for example, pejorative descriptions of Miloš Vystrčil). The social media group, by contrast, relies heavily on reposted content, which signals heightened public interest in single, high-profile topics (*e.g.* the official Czech visit to Taiwan). The news group is comparatively balanced, with varying levels of expressivity and duplication across topics such as sport, business, geolocation, and politics.

The collocation analysis has shown that the most frequent collocates of the noun *Tchaj-wan* ("Taiwan") are primarily associated with political events, international relations, and geolocation terms, including references to Czech political representatives (Miloš Vystrčil, Jaroslav Kubera), visits and delegations, and East Asian countries (China, Hong Kong, Singapore, and Korea). This suggests that the noun is most often used in contexts that frame Taiwan as a political and geopolitical actor situated within broader international and regional debates.

In contrast, the adjective *tchajwanský* ("Taiwanese") frequently co-occurs with company names and economic terminology (*e.g.* TSMC, Foxconn, MediaTek, *investice* "investment"), reinforcing Taiwan's image as a technologically advanced and economically significant partner. This indicates that while *Tchaj-wan* evokes primarily political and diplomatic dimensions, *tchajwanský* reflects economic cooperation and technological development.

Taken together, these findings show that Czech online media between 2017 and 2020 portrayed Taiwan through two complementary lenses: as a democratic political partner and as a technological and economic ally. The representation is predominantly positive, emphasising cooperation, shared democratic values, and mutual benefit. This framing corresponds to the gradual strengthening of Czech–Taiwanese relations during the period in question, particularly in the context of the COVID-19 pandemic and high-level political visits.

The results indicate that the relationship between the Czech Republic and Taiwan is reciprocal, leading to deepened cooperation between 2017 and 2020, a trend that aligns with findings from other, mainly political science, studies (I. T.-Y. Chen 2023; W.-Y. Chen 2023). Nevertheless, the Czech Republic proceeds cautiously, balancing its political relations with both China and Taiwan, despite belonging to a group of countries less dependent on China (Chen 2023: 51). The pandemic demonstrated a further shift of Taiwan towards the Czech Republic. At the same time, Taiwan increasingly appears in Czech media not only as a political partner but also as a technological and economic collaborator, a representation that supports Taiwan's efforts to strengthen its position in the global high-tech supply chain (Chen 2023: 55).

Several limitations should be acknowledged. The study is confined to a specific time period (2017–2020) and focuses exclusively on the Czech language and media environment. Although it includes various online sources, it does not analyse user-generated content in depth, such as reader comments or reposting behaviour, which could provide additional insight into public sentiment. Moreover, as the study sought to identify the most frequent collocations relating to Taiwan, less frequent topics may not have been captured.

Future research could extend this analysis in several directions. A longitudinal approach covering a longer timespan would help identify both stable themes and occasional, time-specific topics. An updated

investigation of the representation of Taiwan in Czech online media could reveal whether ties between the Czech Republic and Taiwan continue to deepen and whether Taiwanese visibility in Czech digital discourse has increased. Comparative studies involving other Central or Eastern European countries could also determine whether similar patterns can be observed elsewhere.

Acknowledgements

References

Baker, Paul (2023) *Using Corpora in Discourse Analysis*. London: Bloomsbury Academic.

Chen, Ian Tsung-Yen (2023) "The Czech Republic's Indo-Pacific Strategy and the Challenges and Opportunities for Taiwan in Europe. [In:] *Taiwan Strategists* No. 7; 37–56.

Chen, Wen-Yu (2023) *Changes in Taiwan-Czech Relations* [thesis, Charles University]. https://dspace.cuni.cz/handle/20.500.11956/186816.Accessed 19.10.2025.

Cvrček, Václav, Zuzana Komrsková, David Lukeš, Petra Poukarová, Anna Řehořková, Adrian Jan Zasina (2021) "From extra- to intratextual characteristics: Charting the space of variation in Czech through MDA." [In:] *Corpus Linguistics and Linguistic Theory* 17(2); 351–382.

Cvrček, Václav, Petr Procházka (2020) *ONLINE1: Monitoring corpus of online Czech*. Ústav Českého národního korpusu FF UK. http://www.korpus.cz. Accessed 19.09.2025.

Elmerot, Irene (2019) "Drunkards and Miniature Mutes: Czech Subjectivity Revealed Through Corpus-Based Discourse Analysis." [In:] *Scando-Slavica*, 65(2); 127–145.

Elmerot, Irene (2025) *Social Group Representation in a Diachronic News Corpus*. Cambridge: Cambridge University Press.

Fidler, Masako (2016) "The Others in the Czech Republic: Their image and their languages." [In:] *International Journal of the Sociology of Language*, 2016(238); 37-58.

Galasiński, Dariusz, Katarzyna Skowronek (2001) "Naming the Nation: A Critical Analysis of Names in Polish Political Discourse." [In:] *Political Communication*, 18(1); 51–66.

Gillings, Mathew, Gerlinde Mautner, Paul Baker (2023) *Corpus-Assisted Discourse Studies*. Cambridge: Cambridge University Press. https://www.cambridge.org/core/elements/corpusassisted-discourse-studies/37A2C82178A5554D9EB2BAF43FBBA74D.

Machálek, Tomáš (2020) "Kontext: Advanced and flexible corpus query interface." [In:] *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille: ELRA; 7003–7008.

Sedláková, Renáta (2015) *Výzkum médií: Nejužívanější metody a techniky*. Praha: Grada Publishing.

Strakhova, Anna (2024) *Korpusová kolokační analýza internetových diskusí na sociální síti Facebook ve vztahu k lidem negroidní rasy, spojená s protesty hnutí 'Black Lives Matter.'* [Bachelor's thesis, Charles University]. https://dspace.cuni.cz/handle/20.500.11956/195073. Accessed 12.09.2025.

Tanaka, Yasuto (2022) "Taiwan and Pacific Command." [In:] Motohiro Tsuchiya, Denny Roy (eds.) *U.S. Indo-Pacific Command: Implications for East Asia* (2022); 121–141.