

Lexical Competition and Change: A Corpus-Assisted Investigation of GAMBLING and GAMING in the Past Centuries

SAGE Open
 July-September 2020: 1–14
 © The Author(s) 2020
 DOI: 10.1177/2158244020951272
journals.sagepub.com/home/sgo


Longxing Li¹ , Chu-Ren Huang² , and Vincent Xian Wang¹ 

Abstract

This article investigates the interplay of lexical competition and socio-historical events through a close examination of the use of GAMBLING and GAMING based on large-scale synchronic and diachronic corpora. We first set the background for comparison through a synchronic study of the collocational patterns and grammatical relations of the two words using Sketch Engine. We show that GAMBLING tends to be associated with negatively perceived activities and strong disapproval, whereas GAMING tends to collocate with recreational activities, business, and technology. Using Google Books Ngram Viewer, we focus on the drastic diachronic changes in use of the two words, from competition to co-development. Based on corpora trends, we correlate the rise and fall of the two words and the change in their competition relation to particular socio-historical events: gold rushes, sports betting, the popularity of video games, and the gaming industry boom. The classical competition model of near synonyms remained valid until recent socio-economic events introduced additional and unique meanings for both words. The article thus shows that linguistic variations as collective human behavior changes can be leveraged to evidence other collective human behavior changes.

Keywords

lexical variations, collective human behavior changes, socio-cultural change, big data linguistics, Gricean maxims

Introduction

This study aims to show that important changes in patterns of collective human behavior can be corroborated and/or discovered by leveraging mature tools to analyze linguistic big data. In particular, we demonstrate that the largest accessible English corpora—enTenTen13 (19 billion words) and Google Books Corpus (361 billion words) (Michel et al., 2011)—are not unyielding but are instead powerful tools for knowledge discovery and linguistic analysis. With the right tools, the linguistic big data will yield information not readily accessible through other approaches. The approach we propose could be characterized as “Big Data Aided Armchair Linguistics” in the spirit of Fillmore (1992).

GAMBLING and GAMING, two words with a close connection and rich implications for socio-economic life and human behavior, have long been the focus of social science research. In addition to the study of gambling- and gaming-related human behaviors, especially in terms of addiction (e.g., Potenza et al., 2019), GAMBLING and GAMING expressions are also studied in discourse studies (Catenaccio, 2015; Haakana & Sorjonen, 2011; Möring, 2013), conceptual metaphor

theory (Lopez-Gonzalez et al., 2017), sociolinguistics (Yan, 2019), translation studies (Pan & Zhang, 2016, 2017), localization (Dong & Mangiron, 2018; Mangiron, 2017; Strong, 2018), and language sources in different discourse types (Ensslin, 2012). The lexical choices in relation to gambling and gaming activities are among the foci of studies. McGowan et al. (2000) compiled an annotated bibliography of the literature in the socio-cultural domain of gaming and gambling, bringing behavioral, text-based, and policy-driven research together to underline the complex links between gaming and gambling. Intriguing questions thus arise: what is the nature of the competition between these two concepts, and are such competition and usage changes reflected in and evidenced by linguistic usage?

¹University of Macau, Macao, China

²The Hong Kong Polytechnic University, Hong Kong, China

Corresponding Author:

Longxing Li, Department of English (E21-2101), Faculty of Arts and Humanities, University of Macau, Macao 999078, China.
 Email: yb67707@um.edu.mo



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of

the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

The *Oxford English Dictionary* (OED, <http://www.oed.com/>) defines GAMING and GAMBLING with the same sense of “the action or practice of playing games, as cards, dice, etc., for stakes.” Yet, GAMING has some other senses that GAMBLING does not have, such as “the action of engaging in games or entertainments; merrymaking; sport” and “the playing of computer (video, etc.) games.” This “partial synonyms” account that restricts the similarity of the two words to one of the senses of GAMING seems straightforward, but does not predict the difference in their usage, such as the use of GAMING to convey positive polarity. Yoong et al. (2013) conducted critical discourse analysis using Fairclough’s three-dimensional framework to deconstruct a Malaysian lottery company’s strategy of promoting lottery activities as “not gambling but gaming.” Pan and Zhang’s (2016, 2017) diachronic studies discovered that the Macao government reframed the gambling business from social problem to entertainment industry through discursive processes in both Chinese and English. Dale (2018) studied the ambiguity of GAMBLING and GAMING in legal discourses using corpus linguistics methods to help lawyers improve their awareness of term ambiguity vis-à-vis their audiences to avoid miscommunication. However, these studies were rather restricted in terms of scope of data and time span. To the best of our knowledge, Li and Huang (2018) is the only large-scale quantitative study on the use of GAMBLING and GAMING, but it relies only on the single synchronic enTenTen13 corpus and its conclusion does not go beyond the distinction of semantic prosody—that is, GAMBLING being more negative while GAMING being more neutral.

In this article, we will investigate three research questions based on the synchronic and diachronic distributional patterns of GAMBLING and GAMING extracted from linguistic big data:

Research Question 1: What is the nature of the co-variation of the meanings of GAMBLING and GAMING?

Research Question 2: What are the co-variation patterns of GAMBLING and GAMING in terms of the changes in collective human behavior?

Research Question 3: Are there regional differences in the co-variation of GAMBLING and GAMING (between American English and British English)?

Method

Our methodology generally belongs to the field of corpus linguistics as an empirical and quantitative approach to studying language in real life (Teubert & Krishnamurthy, 2007). In particular, we focus on identifying widespread patterns of naturally occurring language that may be overlooked by a small-scale analysis (Baker & McEnery, 2005). It is in this spirit that Huang and Yao (2015) viewed corpus linguistics as the precursor to big data linguistics. Corpus linguistic studies rely crucially on both corpus data and analytical

tools. In this study, we adopted Sketch Engine (SkE) developed by Kilgarriff et al. (2004). SkE offers a range of functions, such as Concordancing, Thesaurus, Word Sketch, and Sketch Difference, and has been widely adopted in lexicography, discourse studies, translation studies, language teaching and learning, and so on. SkE can contrast grammatical and collocational relations between near synonyms by processing a huge amount of authentic data (Wang & Huang, 2017) and triangulate the findings obtained through its different functions (Li et al., 2018, 2020). Near synonym-driven research is one of the most productive approaches in corpus linguistics, pioneered by Atkins and Levin (1995) for English and Tsai et al. (1998) for Chinese. By minimizing the lexical contrast, the collocational differences extracted from corpora lead to pinpointed linguistics accounts. In an earlier study using SkE, Li et al. (2018) summarized three groups of grammatical relations (GramRels) from a total of 36 GramRels in SkE—that is, possessive relation, verb–noun relation, and modifying relation—to provide relevant comparisons of three Chinese synonyms. In this study, we selected nine GramRels and summarized them into three groups: coordination relation, verb–noun relation, and modifying relation.

We chose the English web corpus enTenTen13, which contains more than 19 billion words with rich metadata, as our main corpus. The gargantuan corpus size provides comprehensive coverage of a larger variety of linguistic properties. The data were downloaded from the internet in 2013, cleaned, deduplicated, tagged, and spam-filtered (Jakubíček et al., 2013). The last step is vital to our study as GAMBLING and GAMING are among the most popular spam words. The corpus is further divided into sub-corpora according to the top-level domains (e.g., .org, .edu, .com, .uk, and .us) from which the texts are retrieved. This allows users to differentiate between text sources to make comparison across regions.

We also studied the Google Books Corpus to add a historical dimension. The early version of the Google Books Corpus contains 5,195,769 digitized books and more than 500 billion words, with 361 billion in English (Michel et al., 2011). On the free, web-based platform, users can query word usage in one of the languages, or in one of the varieties of English for the period they are interested in. Although the corpus has been criticized, in particular for the quality of the metadata and OCR errors, scholars are still quite optimistic about its future applications (e.g., Nunberg, 2010). This corpus also helps to address Schützler’s (2018) lament about the problem of a lack of diachronic corpora large enough to enable analyses of less-frequent items in diachronic lexical studies. Michel et al. (2011) demonstrated a range of innovative studies in what they called “culturomics”. Twenge et al. (2017) examined trends in the use of seven taboo words in the Google Books Corpus from 1950 to 2008 and found that American culture has become increasingly accepting of the use of taboo words, consistent with higher cultural individualism. Drawing insights from the previous methodologies, our study focuses on the interplay between lexical competition and related socio-historical events.

To study patterns of meaning changes in relation to Research Question 1, we used the Thesaurus function to produce words with a close relation to the keywords. The distributional thesaurus lists words with similar collocational behaviors to the keyword and may include a set of synonyms, antonyms, hypernyms, and hyponyms (Kilgarriff et al., 2014). We hypothesized that these closely related “neighbors” would enable us to identify new meanings and meaning variations. We also examined the extent to which GAMBLING and GAMING are similar to each other with this function and generated common and only patterns through Sketch Difference.

To detect patterns of correlational changes in collective human behavior, as in Research Question 2, we selected three GramRels—that is, “definitions,” “X is a . . .,” and “. . . is a X”—to provide rich descriptive information about the status quo of gambling and gaming in context, with which updated definitions of the terms can be formulated and the latest trends in gambling and gaming activities can be observed. Sentences 1 and 2 are examples of how gaming is represented in the web corpus:

1. Alternate reality gaming is an obsession-inspiring genre that blends real-life treasure hunting, interactive storytelling, video games, and online community and may, incidentally, be one of the most powerful guerrilla marketing mechanisms ever invented.
2. Gaming is a medium that’s closely intertwined with technology.

For the diachronic variations, we utilized three complementary sources—Online Etymology Dictionary, online Oxford English Dictionary, and the Google Ngram Viewer (GNV)—to track the historical changes in the use of GAMBLING and GAMING in terms of meaning and frequency. The Online Etymology Dictionary (<https://www.etymonline.com/>) is a map of the wheel-ruts of modern English, providing explanations of what words meant and how they sounded 600 or 2,000 years ago. The Oxford English Dictionary, containing both synchronic and diachronic information of the English language, provides information to the meaning, history, and pronunciation of 600,000 words from across the English-speaking world. The GNV (<http://books.google.com/ngrams>) provides time sequences of frequency or usages of words and phrases (i.e., *n*-grams) over a period of five centuries in eight languages, covering 6% of all books ever published (Lin et al., 2012). Based on the diachronic patterns extracted from the Google Books Corpus, we attempted to map the changes in important historical events to identify the significant social and cultural changes underlying lexical competition and variations. The Gricean Cooperative Principle (Grice, 1975) and Relevance Theory (Sperber & Wilson, 1986) were drawn on for predictions about the use of GAMBLING and/or GAMING in specific socio-historical contexts. We further mapped GAMBLING and GAMING

Table 1. Thesaurus Lists of GAMBLING and GAMING by Descending Similarity Score.

	Thesaurus
GAMBLING	gaming, poker , casino , <u>trading</u> , <u>smoking</u> , <u>drinking</u> , bingo , <u>addiction</u> , <u>advertising</u> , <u>banking</u> , <u>tourism</u> , <u>abortion</u> , betting , <u>farming</u> , Sport , <u>pornography</u> , hunting , football , <u>fraud</u> , networking , <u>shopping</u> , <u>abuse</u> , blogging , <u>investing</u> , entertainment , <u>selling</u> , driving , <u>crime</u> , <u>theft</u> , <u>promotion</u> , <u>lending</u> , <u>marijuana</u> , . . .
GAMING	gambling, poker , entertainment , networking , <u>computing</u> , photography , <u>advertising</u> , <u>programming</u> , sport , television , <u>publishing</u> , casino , <u>banking</u> , <u>marketing</u> , racing , blogging , learning , football , golf , <u>SEO</u> , <u>manufacturing</u> , fitness , tourism , <u>trading</u> , <u>healthcare</u> , TV , <u>engineering</u> , <u>graphics</u> , <u>promotion</u> , <u>innovation</u> , <u>media</u> , <u>enterprise</u> , . . .

to the Suggested Upper Merged Ontology (SUMO; Niles & Pease, 2001) through WordNet (Pease & Fellbaum, 2010) to understand the conceptual basis for the contrast.

Finally, Research Question 3 regarding regional variation was examined with the Sketch Diff function to sketch and contrast the variations in use of each keyword in the two sub-corpora of enTenTen13—that is, the American English corpus and the British English corpus. The differences were also mapped to possible relevant events in each society.

Results

The results of our analysis are presented here, based on the data and methodology described in the previous section.

Words in the Thesaurus

The 32 words most closely related to GAMBLING and GAMING extracted by SkE thesauri from the enTenTen13 corpus are listed in Table 1. Each of these close “neighbors” can be considered to represent a particular semantic dimension of the keywords. GAMBLING and GAMING are found to be the most closely related word to each other lexically and grammatically (similarity score = 0.376), which reflects their close ties and implies both the difficulty and the necessity of distinguishing one from the other. Closer observation reveals that the thesaurus of GAMBLING can be classified into three major thematic groups: business and industry (dotted underline), recreation (**in bold**), and unhealthy lifestyle or criminal acts (underline). GAMING is mainly associated with industry (dotted underline), recreation (**in bold**), and technology (*in italics*). The thesaurus lists indicate that both GAMBLING and GAMING are associated with industrial and recreational activities, but GAMBLING has a prominent association with unhealthy or even criminal acts (10 of the 32 words), whereas GAMING has an association with technology.

Table 2. Coordination Relation of GAMBLING and GAMING.

“Gambling/gaming” and/or . . .	Counts		Typicality score (LogDice)	
	GAMBLING	GAMING	GAMBLING	GAMING
Prostitution	840	16	8.6	2.6
Sex	710	28	6.1	1.3
Smoking	342	16	6.2	1.6
Alcohol	1,202	54	6.8	2.3
Drinking	1,271	71	8.4	4.1
Pornography	575	52	8.0	4.3
Addiction	581	60	6.9	3.5
Porn	247	49	6.8	4.2
Poker	646	214	7.4	5.7
Casino	1,501	546	8.1	6.5
Lottery	364	233	7.5	6.6
Gambling	1,086	660	8.6	7.6
Betting	275	273	7.3	7.0
Gaming	660	1,252	7.6	8.4
Racing	159	326	5.4	6.3
Entertainment	371	1,386	5.1	6.9
Microgaming	10	194	2.7	6.6
Playtech	9	181	2.6	6.5
Networking	23	471	2.1	6.3
Computing	0	219	—	5.8
Playback	0	140	—	5.9
Anime	0	169	—	6.1
Multimedia	0	277	—	6.7
Browsing	0	379	—	7.2

Similar tendencies can be observed further down the thesaurus list, confirming patterns of association for both GAMBLING and GAMING.

Sketch Diff: Common and Only Patterns

The Sketch Diff function compares two words at a time, generating common patterns of the pair and only patterns of each word. For this study, we compared GAMBLING and GAMING as nouns. We opted for the default setting of a minimal frequency of 10 for both words and a maximum number of items in a GramRel of the common block of 30. The three GramRels examined are as follows: coordination relation, verb–noun relation, and modifying relation.

First, a coordination relation is defined by the context-free pattern of “X and/or” As shown in Table 2, the colors green and red indicate the tendency of listed words to collocate with GAMBLING and GAMING, respectively, in that particular relation. The deeper green a word is, the more likely it is to co-occur with GAMBLING; the deeper red a word is, the more likely it is to co-occur with GAMING. When a word, such as *computing*, does not have an attested “and/or relation” with GAMBLING in the corpus, it is considered to be an only pattern for GAMING. The common patterns and only

patterns in coordination relation provide evidence of semantic differences based on the clustering of similar words. GAMBLING tends to be used in juxtaposition with words associated with an unhealthy lifestyle or even crime (e.g., *prostitution*, *smoking*, *pornography*, *porn*, *addiction*), whereas GAMING is juxtaposed with those related to technology and recreational activities (e.g., *browsing*, *multimedia*, *anime*, *playback*, *computing*). Words that frequently collocate with *both* GAMBLING and GAMING (e.g., *poker*, *casino*, *lottery*, *betting*, *racing*, *entertainment*) in the white area of Table 2 are their common patterns. This indicates the situations in which the two words are to some extent interchangeable. It is worth noting that words in the common patterns still exhibit different degrees of tendency to collocate with the two terms. For example, GAMBLING collocates much more strongly with *casino*, whereas GAMING collocates relatively more with *entertainment*.

Second, we examined verbs that collocate with GAMBLING and GAMING in verb + noun or noun + verb relations, in which GAMBLING and GAMING are either the object or the subject. The verb collocates reveal information about how gambling/gaming activities are carried out, perceived, or regulated by relevant parties. The results automatically generated by SkE inevitably contain occasional noises. Hence, all results were

manually checked to avoid inadvertent errors. It was discovered that while GAMBLING noticeably tends to collocate with verbs related to regulation, policing, or control (e.g., *legalize, counsel, violate, state-sponsor, deduct, criminalize, prohibit, ban, curb, outlaw, and oppose*), GAMING tends to co-occur with verbs pertaining to favorable changes or technological advances (e.g., *redefine, console, revolutionise [ze], innovate, develop, and experience*).

Third, in terms of modifying relation, we focused on both adjectives and nouns that have a modifying relation with GAMBLING and GAMING, including the GramRels of “adjective predicates,” ‘modifiers of gambling/gaming,’ and “nouns modified by gambling/gaming.” The results of the three GramRels are shown in Tables 3 to 5. Again, on one hand, gambling is found in collocation with adjectives with a strong negative meaning (e.g., *immoral, sinful, unlawful, risky, regressive, unregulated, rampant, evil, illegal, pathological*). On the other hand, GAMING collocates with adjectives with a neutral or positive meaning, indicating its pervasiveness and predominance (e.g., *massive, public, booming, populous, mainstream, ubiquitous*) as well as its technological (e.g., *real time, PC, 3D, etc.*) and recreational (e.g., *interactive, pleasurable, sensible, casual*) nature. The nouns modifying and being modified by gambling can be classified into three main themes—that is, venues or facilities (e.g., *den, roulette, mecca, establishment, sites*), gambling-related actions or activities (e.g., *drinking, bankroll, bet, game, bonus*), and (mostly negative) consequences (e.g., *debt, hell, winning, monopoly, addict, addiction*). By contrast, nouns strongly associated with GAMING tend to be either technical/technological terms (e.g., *console, 3D, retro, PC, laptop, video, software, platform, headset, tabletop*) or the proper names of large gaming companies or organizations (e.g., *Boyd, Cantor, SK, WMS*).

Representation of GAMBLING and GAMING in Activities

Tables 6 and 7 present information about how GAMBLING and gaming activities are represented through the three GramRels. To focus on the distinctness of each activity, we introduce the only patterns first in Table 6, summarizing the only patterns of gambling and GAMING for the three GramRels, with the words sorted in descending order of salience. Table 7 illustrates the common patterns with the frequency of collocation in three GramRels for each keyword. Yet again, GAMBLING occurs with nouns that denote very negative evaluations—for example, *addiction, risk, disorder, problem, and sin*—evoking clearly negative representations of the activity involved, although it also collocates with a range of generic umbrella terms such as *behavior, act, phenomenon, practice, and activity*. By contrast, the collocates of GAMING mainly pertain to the participants and stakeholders in the gaming industry—for example, *developer, provider, market, company, iNetBet*—or the platforms, tools, and types of gaming—for example,

Table 3. Adjective Predicates of GAMBLING and GAMING.

Adjective predicates	Counts		Typicality score (LogDice)	
	GAMBLING	GAMING	GAMBLING	GAMING
Immoral	31	0	6.0	—
Sinful	25	0	5.8	—
Unlawful	21	0	5.5	—
Risky	42	0	5.3	—
Regressive	10	0	5.1	—
Lawful	14	0	5.1	—
Synonymous	43	0	5.1	—
Unregulated	9	0	4.7	—
Widespread	33	0	4.7	—
Lucrative	11	0	4.6	—
Joint	20	0	4.6	—
Rampant	18	0	4.4	—
Evil	19	0	3.8	—
Illegal	1,070	66	8.7	4.7
Legal	938	115	8.6	5.6
Prevalent	35	11	4.2	2.6
Addictive	110	38	7.1	5.7
Online	940	426	4.0	2.9
Convenient	35	17	3.9	2.9
Entertaining	24	21	4.3	4.2
On-line	27	29	3.7	3.8
Popular	203	228	4.4	4.6
Exciting	50	59	4.0	4.2
Social	11	28	2.5	3.9
Enjoyable	20	57	3.1	4.7
Massive	0	13	—	3.8
Sensible	0	9	—	4.0
Public	0	18	—	4.0
Gross	0	10	—	4.0
Booming	0	14	—	4.1
Interactive	0	15	—	4.2
Concerned	0	20	—	4.2
Pleasurable	0	10	—	4.4
Populous	0	9	—	5.3
Animate	0	9	—	5.5
Mainstream	0	14	—	5.5
Ubiquitous	0	26	—	5.6
Demographic	0	44	—	6.1
Peripheral	0	50	—	7.4

software, medium, genre, category, live, console, site—conveying a neutral image.

The common patterns of GAMBLING and GAMING summarized in Table 7 show that both words are most frequently represented as a *game, activity, business, industry, hobby, or pastime*. They also have a similar number of instances of being described as a *sector* or an *issue*. However, notable differences in representation are still evident through these common patterns. For example, GAMBLING is much more

Table 4. Modifiers of GAMBLING and GAMING.

Modifiers	Counts		Typicality score (LogDice)	
	GAMBLING	GAMING	GAMBLING	GAMING
Prostitution	177	0	5.4	—
Problem	3,554	90	7.6	2.2
Pathological	1,079	45	7.9	2.9
Compulsive	1,260	56	8.1	3.2
Pornography	219	14	5.7	1.3
Illegal	1,954	125	6.8	2.8
Unlawful	316	24	5.9	1.8
Underage	405	32	6.5	2.4
Riverboat	178	27	5.6	2.4
Regulated	283	89	5.8	3.8
Roulette	516	167	6.5	4.5
Internet	6,399	2,356	6.5	5.0
Expanded	388	203	5.8	4.6
Blackjack	282	158	5.7	4.5
Casino	9,875	4,812	8.8	7.7
Gambling	413	219	5.4	4.3
Online	29,103	17,577	7.4	6.7
Real-money	196	243	5.7	5.6
Responsible	729	755	6.0	5.9
On-line	495	539	5.7	5.6
Mobile	863	4,180	3.8	6.1
Real	120	1,282	3.2	6.5
Video	285	6,817	2.8	7.4
Multiplayer	34	1,130	2.7	7.4
Casual	31	987	1.3	6.1
Realtime	15	769	1.9	7.1
Pc	15	3,542	—	7.9
Handheld	0	447	—	6.0
Retro	0	514	—	6.0
Cantor	0	376	—	6.2
Tabletop	0	409	—	6.2
3d	0	1,516	—	6.3
Sk	0	460	—	6.4
Boyd	0	507	—	6.5
Rival	0	710	—	6.8
Console	0	1,494	—	7.7

strongly represented as an *addiction* ($n = 128$) than GAMING is ($n = 17$), whereas GAMING is far more frequently portrayed as a *hobby* ($n = 210$) than GAMBLING is ($n = 59$). The fine-grained differences match and support our earlier generalization regarding the polarity differences of these two near synonyms.

Regional Variation

In this section, we focused on regional variations in the use of GAMBLING and GAMING in the two major varieties of English: American English and British English. GAMBLING shows

Table 5. Nouns Modified by “GAMBLING/GAMING.”

Nouns modified by	Counts		Typicality score (LogDice)	
	GAMBLING	GAMING	GAMBLING	GAMING
Den	3,318	96	8.8	2.9
Debt	2,494	53	6.2	0.4
Hell	1,257	45	7.2	1.6
Winning	687	36	6.6	1.5
Establishment	4,898	756	8.3	5.1
Addict	1,039	243	6.8	4.0
Addiction	3,212	876	8.0	5.6
Casino	5,714	1,717	7.5	5.5
House	7,019	2,359	6.0	4.4
Site	11,709	9,082	5.8	5.4
Operator	1,511	1,468	5.9	5.6
Portal	820	1,091	5.8	5.7
Commission	2,654	3,719	5.9	6.2
Revenue	1,103	2,800	5.4	6.5
Industry	6,134	16,543	5.8	7.1
Machine	1,203	5,105	4.3	6.3
Table	735	4,310	3.9	6.3
Experience	3,155	22,553	4.1	6.9
Platform	273	5,898	2.4	6.6
Headset	12	1,415	0.6	6.7
Laptop	21	4,436	0.7	7.9
Console	23	13,489	0.8	9.5
Rig	0	1,307	—	6.4
Mouse	0	2,682	—	7.2
Pc	0	3,167	—	7.3

notably different distributions in coordination relation, noun–verb relation, and, in particular, modifying relation between American English and British English. Because the relative sizes of the U.K. (roughly 1,182 million words) and the U.S. (164 million words) corpora strongly favor U.K. usages, which could lead to “over-magnified” U.K.-only patterns, our discussion here focuses on the U.S. only patterns.

One salient group identified as U.S.-specific modifiers of GAMBLING are the names of American football teams—for example, *Seahawks*, *Texans*, and *Ravens* (Table 8). While this is not a surprising finding, it is remarkable that no U.K.-specific team names (e.g., Manchester United) occur in the only patterns. This indicates the popularity of betting on football games in the U.S. (cf. sports betting in the “Discussion” section). Nouns modified by GAMBLING reveal similar differences between British and American English.

There are also clear differences between British and American English in the modifiers of GAMING (Table 8). The predominant modifiers of GAMING in the United States tend to be sports-related, such as *football*, *rugby*, *hockey*, *alternate*, *preseason*, *playoff*, and *NFC* (National Football Conference). This again consolidates the above findings in relation to the

Table 6. Only Patterns of GAMBLING and GAMING in Three GramRels.

GramRels	GAMBLING	GAMING
Definitions	gambling, addiction, leisure, behavior, act, phenomenon, risk, disorder, issue, sport, problem, practice	gaming, multi-billion, dollar, developer, software, provider, market, medium, company
X is a . . .	gambling, sin, poker, behavior, luck, lottery, vegas, disorder, habit, tax, profession, risk	anomaly, gaming, medium, genre, future, category, escape, developer, provider, interest
. . . is a X	gambling, statute, betting, poker, roulette, trading, activity, option, people	iNetBet, live, gaming, console, focus, piece, site

Table 7. Frequency of Common Patterns of GAMBLING and GAMING in Three GramRels.

Representation	Definitions		X is a is a X		Total	
	GAMBLING	GAMING	GAMBLING	GAMING	GAMBLING	GAMING	GAMBLING	GAMING
Game	116	21	269	67	20	32	405	120
Activity	79	25	262	107			341	132
Fun	93	49					93	49
Entertainment	22	12	72	17			94	29
Pastime	29	18	81	59			110	77
Business	57	45	160	99			217	144
Trend	9	9					9	9
Industry	33	59	90	130			123	189
Casino	19	41	57	45	21	45	97	131
Hobby	9	63	50	147			59	210
Addiction			128	17			128	17
Pursuit			17	12			17	12
Sector			12	18			12	18
Market			24	48			24	48
Passion			12	43			12	43
Issue					17	13	17	13

Table 8. Only Patterns of “Modifiers of GAMBLING/GAMING” in the U.S. Sub-Corpus.

Modifiers of GAMBLING	Frequency	Score	Modifiers of GAMBLING	Frequency	Score
Seahawks	7	4.0	Alternate	9	4.5
Football	7	4.0	Football	6	3.8
Texans	6	4.0	Preseason	13	3.5
Ravens	6	3.2	Playoff	18	2.2
Table	103	2.8	NFC	7	2.2
Bet	11	2.7	Rugby	6	0.9
Fatigue	13	2.5	Hockey	8	0.4
Winner	12	1.5			
Cyber	14	1.2			

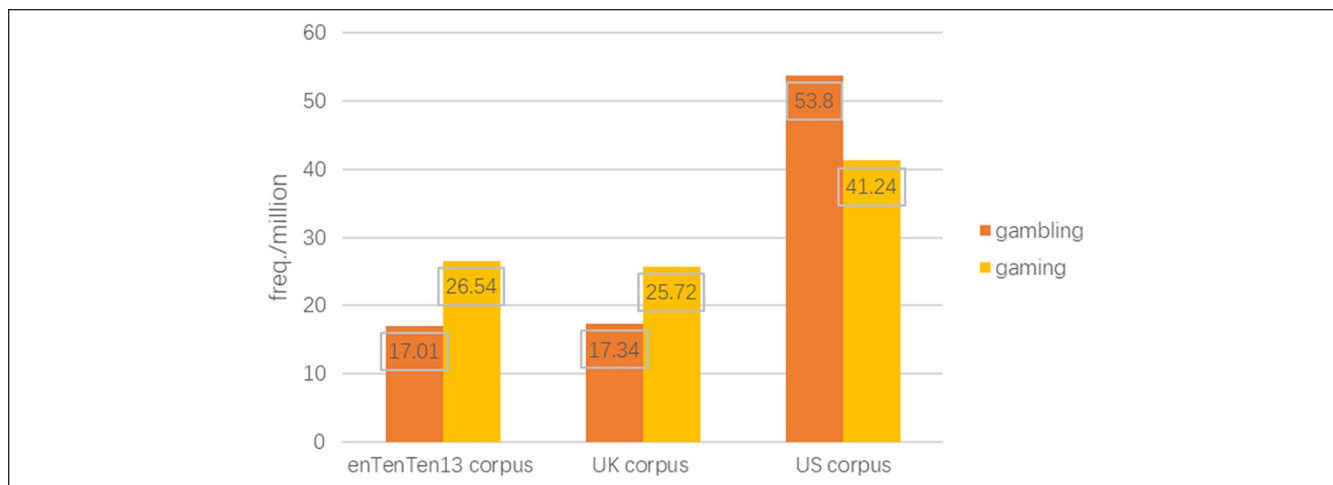
modifiers of GAMBLING, that sports betting tends to be a common practice in the United States. Apart from the differences in collocations, there are regional differences in terms of the two words' frequency of occurrence (see Table 9).

Taking the entire enTenTen13 corpus as a reference, the normalized frequency of both words in the U.K. sub-corpus is very close to that in the corpus as a whole, whereas that of GAMBLING and GAMING in the U.S. sub-corpus is more than 3.1 and 1.5 times higher, respectively, than in the enTenTen13

corpus as a whole (Figure 1). This indicates that both GAMBLING and GAMING are used much more frequently in the United States, most probably reflecting the prevalence of GAMBLING and gaming activities there. We used UCREL log-likelihood wizard (Rayson & Garside, 2000) to perform tests for a significant difference in the frequency of each word between the U.K. and the U.S. sub-corpora. A log-likelihood of 6.6 or above illustrates that the difference is significant at $p < .01$ level. The log-likelihood scores show that the

Table 9. Frequency of Occurrence of GAMBLING and GAMING in enTenTen13.

Word	U.K. (1,182,251,470)	U.S. (164,190,640)	enTenTen13 (19 billion)
GAMBLING	20,498	8,834	386,782
GAMING	30,402	6,772	603,436

**Figure 1.** Normalized frequency of GAMBLING and GAMING in enTenTen13.

frequency differences between the U.K. corpus and the U.S. corpus for both GAMBLING and GAMING are significant. In addition, GAMBLING is more frequently used than GAMING in the United States, whereas GAMING is more commonly used in the United Kingdom and the enTenTen13 corpus as a whole.

Diachronic Change

We then used dictionaries and GNV to track diachronic changes in the use of GAMBLING and GAMING in terms of meaning and frequency. The Online Etymology Dictionary provides a brief historical account:

gambling: (n.) 1784, “habitual indulgence in gambling,” verbal noun from gamble (v.). Gambling-house attested by 1794.

gaming: (n.) 1500, “gambling,” verbal noun from game (v.). From 1980s in reference to video and computer games. Gaming-house is from 1620s; gaming-table from 1590s.

OED traced GAMBLING a little further back to 1700 in the quotation “The Room where it stood was an old gambling Cock-loft” from a translation related to the Spanish novel *Don Quixote*. OED’s earliest record of GAMING with the meaning “the action or practice of playing games, as cards, dice, etc., for stakes” is in 1501, consistent with the finding in the Etymology Dictionary, although the word is spelled as *gamyng*. In addition, OED reveals that GAMING and GAMBLING

are etymologically linked by the obsolete word *gameling* (earliest instance 1594), which means “the playing of games; (perhaps) gambling.” *Gamble*, the verb form of GAMBLING (1700), first occurred in the 1750s. Based on this information, we can confirm that both GAMBLING and GAMING are derived from “game,” following the sequence below:

game, v. → *gaming*, n. → *gameling*, n. → *gambling*, n. → *gamble*, v.

We queried GNV for the occurrence of GAMBLING and GAMING in the Google Books English corpus (Figure 2). Before the 1750s, GAMING has achieved sizable usage while GAMBLING was rarely used. The frequency of use of both words fluctuated noticeably in the early stages due to the scarcity and uneven distribution of historical data, so we focus on trends after the 1750s, and divide them into three periods. In the first period (1750–1900), the use of GAMING increased rapidly and remained at a high level until the 1820s, but steadily declined toward the end of the 19th century. By contrast, the use of GAMBLING quickly rose from a low-level, surpassed GAMING in 1847 and continued to increase (cf. The Gold Rush era in the “Discussion” section). In the second period (1900–1980s), GAMBLING maintained a frequency around three times as high as that of GAMING, whereas the First World War (WWI) and the Second World War (WWII) saw temporary downturns in the use of both words. Finally, from the early 1990s, use of the two words rose sharply in parallel, although GAMBLING

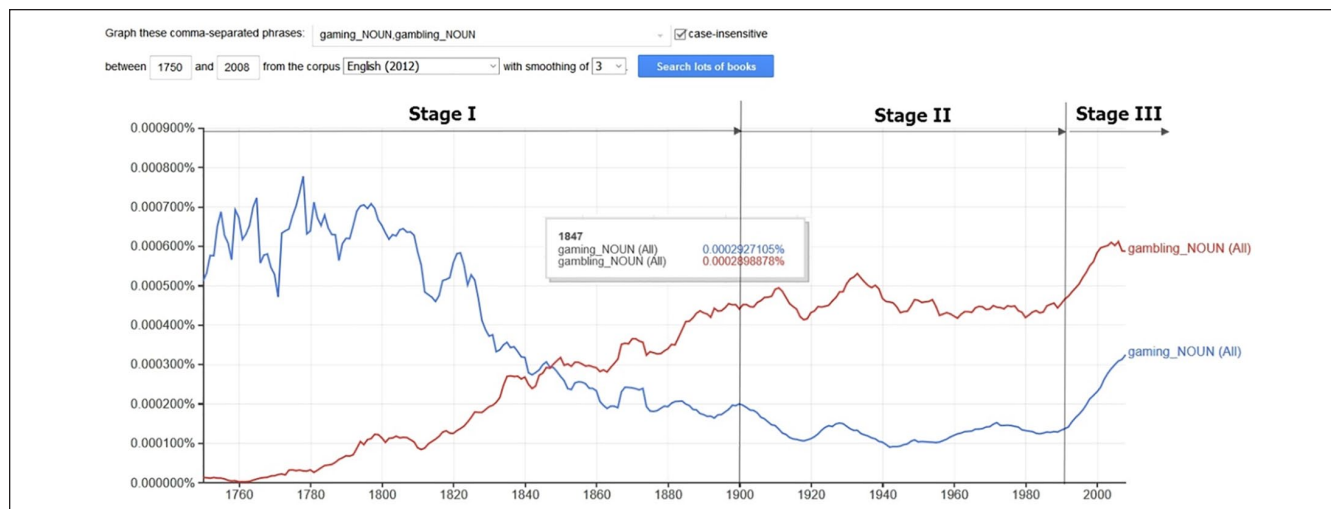


Figure 2. Google Ngram Viewer of GAMBLING and GAMING (1750–2008).

was always more frequently used than GAMING by a wide margin (cf. elaboration on socio-historical factors in the “Discussion” section).

The trends exhibited in the overall English corpus closely match the trends for both the American and the British English sub-corpora in GNV, although GAMBLING occurred at markedly higher frequencies in American English than in British English over the past 150 years, particularly from the 1990s onwards.

Discussion

Based on the above results, we now discuss some key socio-historical events and factors that have contributed to the rise and fall in usage of the two words over the last two centuries. We draw on the Gricean Cooperative Principle and Relevance Theory to pinpoint the pragmatic motivations for preferring one word over the other, and relate the word usage to the relevant conceptual frameworks, based on the Ontology–Lexicon Interface (OntoLex, Huang et al., 2010). We argue that the dominant model that captures the relations of the two words shifts from the competition model in the 19th century to the co-development model from the 1990s.

The Gold Rush Era: Rise of GAMBLING and Decline of GAMING

The increase in the use of GAMBLING and the decline in use of GAMING largely coincided with the “gold rush” period in the English-speaking world, peaking around the years of the California Gold Rush (1848–1855). Gambling in gold-mining areas has been recorded in English-speaking countries such as Australia, New Zealand, South Africa, the United States, and Canada (cf. Courtwright, 1996; Fetherling & Fetherling, 1997). In the United States, the years following

the discovery of gold in California (1848) witnessed the influx of gold miners and the sprouting of saloons and gambling halls “everywhere” (Courtwright, 1996, p. 71). Courtwright noted that gambling was part of the gold rusher’s life—“[i]n such a risk-all atmosphere, with death and illness daily prospects, gambling was a natural pastime” (Courtwright, 1996, p. 72). In the eyes of one San Franciscan, “[e]verybody gambled [. . .] that was the excuse for everybody else” (Courtwright, 1996, p. 73).

GNV captures moderate use of “gold rush” lemma in American English around the U.S. gold rush era, along with much more frequently occurring alternative expressions such as “gold discovery” and “gold mining.” The continued increase in the occurrence of the “betting_NOUN” lemma was also noticeable in this period. Gold rushes emerged as a socio-historical event in which the practice of gambling and (money) betting reached a considerable volume, and the word GAMBLING gained the momentum for a steady rise in usage. The frequency of occurrence of GAMBLING quadrupled in both American and British English in the 19th century, whereas the use of GAMING declined by more than two thirds in both varieties of English (cf. Figure 2). Based on the trends and timing, the gold rush is likely to have been the catalyst among other events favoring the popularity of gambling.

The dominance of GAMBLING at this stage can be predicted by both the Gricean sub-maxim of quantity—that is, “make your contribution as informative as is required (for the current purposes of the exchange)” (Grice, 1975, p. 45)—and the principle of maximizing the relevance to cognition in Relevance Theory (Wilson & Sperber, 2006). GAMBLING entails sufficient and precise enough information to denote betting-for-money activities, whereas GAMING sounds vaguer and is less effective in achieving the communicative purpose. The former therefore eventually overtook the latter around the mid-19th century.

The Surge of GAMING: The 1980s Onwards

Since the late 1980s, both GAMBLING and GAMING have shown a sharp and continuous rise in frequency count (cf. Results). We discuss the reasons for the sharp rise in the use of GAMING in this section and that of GAMBLING in section “GAMBLING and Sports in the United States.” We found two factors underlying the recent surge in the use of GAMING—(a) the desire of the industry and government to switch from GAMBLING to GAMING in their discourses and (b) the substantial spread of (computer-based) video games from the late 1980s.

First, recent studies (e.g., Pan & Zhang, 2016; Yoong et al., 2013) have revealed that the gambling industry and governments tend to shift their usage from GAMBLING to GAMING when they refer to the industry in an attempt to change the public’s perception and reframe the socially problematic business as an important source of revenue. GAMING is more favorable than GAMBLING in terms of its semantic and conceptual properties. GAMING tends to be represented as a *market*, a *company*, a *medium*, or a *genre*, whereas GAMBLING often collocates with words with a negative meaning—for example, *addiction*, *risk*, *disorder*, *problem*, or *sin* (cf. Results). The SUMO network reveals more fine-grained differences. SUMO (Niles & Pease, 2001; Pease, 2011), the only formal ontology that has been mapped to all of the approximately 117,000 word senses in WordNet lexicons (Fellbaum, 1998), identifies the lexicalized conceptual differences between GAMING and GAMBLING, although both words are considered to be complex events. According to SUMO, gaming is both a *Contest* and a *Recreation or Exercise*, which is defined as “A Contest whose purpose is the enjoyment/stimulation of the participants or spectators of the Game.” By contrast, gambling is a *Game* and *Betting* at the same time, and *Betting* is defined as “A Financial Transaction where an instance of Currency Measure is exchanged for the possibility of winning a larger instance of Currency Measure within the context of some sort of Game.” Since gambling is a sub-type of gaming, we can observe that shifting from using GAMBLING to GAMING serves to de-emphasize the characteristics inherent in *Betting* while embracing the characteristics that are conceptually closer to GAMING. In terms of communication, this preference is an attempt to “optimize” the meaning by strengthening the positive meaning and dissociating from the negative meaning (Sperber & Wilson, 1986).

Second, the emergence of (computer-based) video games created strong demand for the use of GAMING. Video games debuted in the mid-20th century and were commercialized in the early 1970s, but also experienced several crashes and recessions (Wolf, 2012)—for example, the great crash in North America in 1983. It is generally believed that the industry was almost single-handedly revitalized by the Japanese corporation Nintendo in 1985, and video gaming culture has become well established since the widespread success of the Nintendo Entertainment System in the United States, followed by Europe, Australia, and other regions

(Consalvo, 2006). Since the 1980s, video gaming has become a major form of entertainment, especially for the young. Figure 3 shows that the “gaming_NOUN” lemma increased in frequency from the late 1980s at a speed similar to that of “video games” and “computer games,” suggesting a strong positive correlation between the spread of video games and the increase in the use of the word GAMING. GAMING provides sufficient and precise information about the playing of video or computer games, satisfying the Gricean maxim of quantity and Relevance Theory of maximizing meaning for cognition. In addition, based on SUMO conceptual ontology, GAMING is not only a synonymous alternative and euphemism for GAMBLING but also a hyponym of video gaming and wagering activities, suited to a much wider range of usage domains.

GAMBLING and Sports in the United States

GAMBLING tended to be used more frequently in American English than in British English over the past 150 years in the sub-corpora in GBV and also more recently in enTenTen13 (cf. Results). We would argue that this reflects the practice of sports betting in the United States (cf. Figure 4). The popularity of betting on sports in the United States has been clearly documented in gambling studies—for example, Thompson (2015) observed that “making wagers on the results of games may be the most popular form of gambling in North America and it is certainly the most popular form of illegal betting in the US” (p. 387). We also found that the most salient usage of GAMBLING in American English occurs in the context of (professional) sports (cf. the “Regional Variation” section).

Interestingly, sports betting was actually prohibited in the United States by the Professional and Amateur Sports Protection Act of 1992, except in four states that had already legalized it—Delaware, Montana, Nevada, and Oregon (Dayanim et al., 2018). The reason why the use of GAMBLING related to sports increased at the same time as it became illegal in most states lies in the fact that gambling activities still took place outside of the four states, and the U.S. government and the press wanted to underline the illegality of these activities and differentiate them from GAMING. In other words, the use of GAMBLING to specifically refer to sports betting satisfied the need to maximize meaning by underlining its illegal (in most cases) status while satisfying the communication need to optimize meaning by allowing GAMING to refer to sports that did not involve betting. In terms of Gricean maxims, it is similarly the competition between the maxim of relevance and the supermaxim of ‘being perspicuous’—that is, avoiding ambiguity (Grice, 1975, p. 46). This accounts for the rise in use of both words in parallel from the late 1980s in American English. Here we must emphasize that such correlations were surmised and corroborated in our research, but not proven. For instance, we suspect that the formation and growing power of governing bodies of sports in the mid-1990s (Forster, 2016) may also

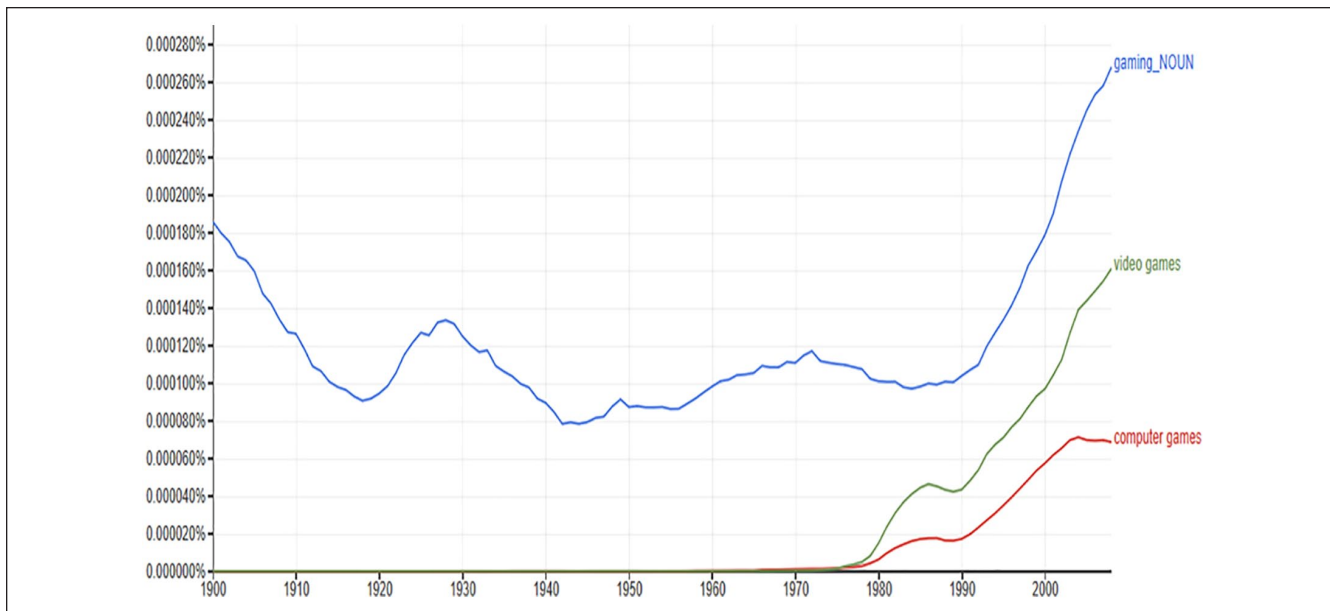


Figure 3. “Gambling,” “video games,” and “computer games” in English (1900–2008, GNV).
 Note. GNV = Google Ngram Viewer.

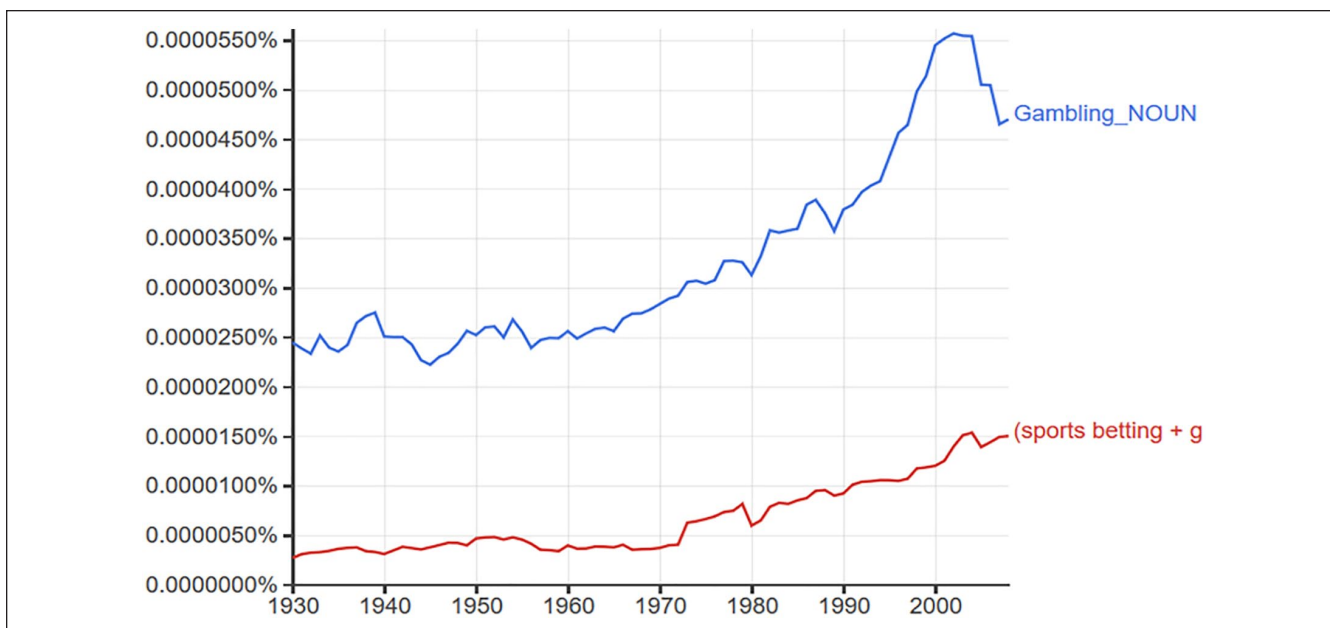


Figure 4. “Gambling” and “sports betting” words in American English (1930–2008, GNV).
 Note. GNV = Google Ngram Viewer.

have contributed to the increased usage of GAMING, but our data simply do not indicate a strong enough pattern to support this hypothesis.

In terms of semantic prosody, we observed that the widely used sports-related collocations of GAMBLING in American English largely have the effect of mitigating the harms and problems associated with gambling activities while highlighting the interactional and recreational features of sports

as an integral element of gambling (cf. Results). By contrast, the largely negative terms collocated with GAMBLING in the U.K. corpus probably reflect its narrower usage in British English.

An explanation for the greater frequency of occurrence of GAMBLING in the United States than in the United Kingdom can also be constructed based on ontology. Sports-related terms, as featured collocations of GAMBLING in the U.S.

corpus, tend to meet the desirable conceptual associations of participants in or spectators of games, including enjoyment and stimulation as well as recreation or exercise. In this favorable view of sports, gambling on professional sports tends to be considered more socially acceptable. In fact, sports betting continued to grow regardless of the legal constraints in the United States and was finally legalized by federal law in 2018.

The Competition Model and the Co-Development Model

The chronological data from this study provide evidence not only for the competition between GAMBLING and GAMING in earlier times but also for their rapid rise in parallel in recent decades. This evokes two different models.

The fact that GAMBLING and GAMING are near synonyms, although with hyponymic relation, suggests that they should be in competition. The earlier trend in usage bore out this prediction—GAMBLING exhibited a steady rise in usage from the late 18th century into the early 20th century, whereas GAMING was in continuing decline. The sum of their frequencies was roughly stable, suggesting that GAMBLING, the newcomer, was to a large extent replacing GAMING in usage. We can observe that there are bi-grams in which GAMING was gradually replaced by GAMBLING, leading to the shift, for example, from “gaming debts” to “gambling debts” and from “gaming houses” to “gambling houses,” in around the 1840s. The competition model prevailed in this period.

However, from the 1990s, both GAMBLING and GAMING increased exponentially in frequency of use. A new model emerges, which we named the “co-development model.” The co-development period of the two words corresponds to a time that saw the opening up and expansion of gaming businesses on a transcontinental scale, from North America to Europe and the Asia-Pacific region (Thompson, 2009). In the United States, given the staggering “underground” sports gambling market, there has been extensive debate, in official settings and on the web, surrounding the voter referenda and the legalization of sports betting in 2018. The booming industry across the world, including in Macao, Singapore, Australia, Canada, South Korea, and the United Kingdom, created an unprecedented need to use the words GAMBLING and GAMING in both professional and mass communication.

Apart from the (external) social factors, we hold that, linguistically, the co-development of GAMBLING and GAMING is possible only if the two words stand as distinct lexical options, each uniquely suitable for certain communicative settings and purposes. These words are therefore not fully interchangeable and are both needed to communicate gambling/gaming matters. We have observed that GAMBLING and GAMING are complementary in terms of their functions in the context of sports betting in the United States. Therefore, with the recent boom in the gaming industry, GAMBLING and

GAMING co-existed and greatly increased in their frequency of occurrence, exhibiting a shift from the competition model to a co-development model, each with different additional semantic functions—that is, video games for GAMING and sports betting for GAMBLING.

Conclusion

This study has used two of the largest language data sets (enTenTen13 and Google Books Corpus) with SkE and GNV to examine GAMBLING and GAMING. It is found that GAMBLING tends to have a more negative semantic prosody, whereas GAMING is more neutral. The OntoLex approach further casts light on their conceptual differences. Diachronically, three periods emerged—from two early periods in competition that resulted in shifting dominance from GAMING to GAMBLING, to recent decades when usage of both terms rose in parallel. The competition of cognitive and communication needs of Relevance Theory (Sperber & Wilson, 1986; Wilson & Sperber, 2006) nicely accounts for the dynamic co-variation of the two terms, as did the application of different Gricean maxims. The changes in the pattern of co-variations occurred in the context of possible socio-economic catalysts such as the mid-19th century gold rushes, and the late-1990s success of video-game products, as well as the popularity and semi-legalization of sports betting in the United States. The shift from the competition model of near synonyms to the unusual co-development model seems to be the result of a coincidental simultaneous surge in the popularity of video games and sports betting. That is, these surges created semantic space for GAMING and GAMBLING, respectively, freeing them from the paired competition relation of the past 300 years.

To conclude, we wish to draw attention with this study not only to the potential of linguistic big data-driven research but also to the simple fact that any linguistic change is a change in collective human behavior. Observation, explanation, and prediction of collective human behavior changes might be the single most impactful research issue in the humanities and social sciences. Yet, historical changes in collective human behavior are extremely difficult to pin down. However, based on the self-evident truth that changes are either a reaction to or instigated by other changes, we were able to use the historical usage record of a pair of near synonyms that represent a significant human behavior to surmise and later corroborate potential major changes. The co-dependency of gambling, sports, and gaming is in fact well documented (Macey & Hamari, 2018). This ability to provide evidence for changes in collective human behavior may well be the most valuable contribution of linguistic big data.

Acknowledgments

Longxing Li and Vincent X. Wang would like to acknowledge the research project MYRG2019-00162-FAH of the University of Macau, which provided the PhD scholarship for Longxing Li.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was supported by the MYRG2019-00162-FAH of the University of Macau and the Joint Supervision Scheme with the Chinese Mainland, Taiwan, and Macao Universities sponsored by the Hong Kong Polytechnic University (Project No. G-SB97).

ORCID iDs

Longxing Li  <https://orcid.org/0000-0002-0532-7956>

Chu-Ren Huang  <https://orcid.org/0000-0002-8526-5520>

Vincent Xian Wang  <https://orcid.org/0000-0002-0203-1973>

References

- Atkins, B. T., & Levin, B. (1995). Building on a corpus: A linguistic and lexicographical look at some near-synonyms. *International Journal of Lexicography*, 8(2), 85–114.
- Baker, P., & McEnery, T. (2005). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics*, 4(2), 197–226.
- Catenaccio, P. (2015). The representation of gambling in gambling awareness campaigns: The discursive construction of addiction. In M. Gotti, S. M. Maci, & M. Sala (Eds.), *Insights into medical communication* (pp. 269–292). Peter Lang CH.
- Consalvo, M. (2006). Console video games and global corporations: Creating a hybrid culture. *New Media & Society*, 8(1), 117–137.
- Courtwright, D. T. (1996). *Violent land: Single men and social disorder from the frontier to the inner city*. Harvard University Press.
- Dale, K. (2018). Legal corpus linguistics: Gambling to gaming language powers and probabilities. *UNLV Gaming Law Journal*, 8(2), 233–252.
- Dayanim, B., Jacob, R., & Georg, E. (2018). Overview of US Federal Gaming Law. In C. Rohsler (Ed.), *The Gambling Law review* (pp. 23–41). Law Business Research Ltd.
- Dong, L., & Mangiron, C. (2018). Journey to the East: Cultural adaptation of video games for the Chinese market. *Journal of Specialised Translation*, 29, 149–168.
- Ensslin, A. (2012). *The language of gaming*. Palgrave Macmillan.
- Fellbaum, C. (1998). *WordNet: An electronic lexical database*. MIT Press.
- Fetherling, D., & Fetherling, G. (1997). *The gold crusades: A social history of gold rushes, 1849-1929*. University of Toronto Press.
- Fillmore, C. J. (1992). “Corpus linguistics” or “computer-aided armchair linguistics.” In J. Svartvik (Ed.), *Directions in corpus linguistics: Proceedings of Nobel symposium 82* (pp. 35–60). De Gruyter Mouton.
- Forster, J. (2016). Global sports governance and corruption. *Palgrave Communications*, 2, Article 15048. <https://doi.org/10.1057/palcomms.2015.48>
- Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). Academic Press.
- Haakana, M., & Sorjonen, M. L. (2011). Invoking another context: Playfulness in buying lottery tickets at convenience stores. *Journal of Pragmatics*, 43(5), 1288–1302.
- Huang, C. R., Calzolari, N., Gangemi, A., Lenci, A., Oltramari, A., & Prevot, L. (Eds.). (2010). *Ontology and the lexicon: A natural language processing perspective*. Cambridge University Press.
- Huang, C. R., & Yao, Y. (2015). Corpus linguistics. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 949–953). Elsevier.
- Jakubiček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The tenten corpus family. In A. Hardie & R. Love (Eds.), *7th International corpus linguistics conference* (pp. 125–127). Lancaster: UCREL.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36.
- Kilgarriff, A., Rychlý, P., Smrz, P., & David, T. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), *Proceedings of the 11th EURALEX international congress* (pp. 105–116). Lorient, France: Université de Bretagne-Sud.
- Li, L. X., Dong, S. C., & Wang, V. X. (2020). Gaige and reform: A Chinese-English comparative keywords study. In Q. Su & W. Zhan (Eds.), *From minimal contrast to meaning construct: Corpus-based, near synonym driven approaches to Chinese lexical semantics* (pp. 321–332). Springer.
- Li, L. X., & Huang, C. R. (2018). *Discriminating near synonyms with Sketch Engine: A comparative study of gambling and gaming* [Paper presentation]. SNU International Conference on Linguistics, Seoul.
- Li, L. X., Huang, C. R., & Gao, X. F. (2018). A SkE-assisted comparison of three “prestige” near synonyms in Chinese. In J. F. Hong, Q. Su, & J. S. Wu (Eds.), *Chinese lexical semantics. CLSW 2018. Lecture notes in computer science* (Vol. 11173 pp. 256–266). Springer.
- Lin, Y., Michel, J.-B., Aiden, E. L., Orwant, J., Brockman, W., & Petrov, S. (2012, July). *Syntactic annotations for the Google Books Ngram corpus*. Proceedings of the ACL 2012 System Demonstrations, Jeju Island, Korea.
- Lopez-Gonzalez, H., Guerrero-Solé, F., Estévez, A., & Griffiths, M. (2017). Betting is loving and bettors are predators: A conceptual metaphor approach to online sports betting advertising. *Journal of Gambling Studies*, 34(2), 1–18.
- Macey, J., & Hamari, J. (2018). Investigating relationships between video gaming, spectating esports, and gambling. *Computers in Human Behavior*, 80, 344–353.
- Mangiron, C. (2017). Research in game localisation: An overview. *The Journal of Internationalization and Localization*, 4(2), 74–99.
- McGowan, V., Droessler, J., Nixon, G., & Grimshaw, M. (2000). *Recent research in the socio-cultural domain of gaming and gambling: An annotated bibliography and critical overview*. Alberta Gaming Research Institute.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176–182.

- Möring, S. (2013.). *Games and metaphor—A critical analysis of the metaphor discourse in game studies* [Ph.D. thesis]. IT University Copenhagen.
- Niles, I., & Pease, A. (2001). Towards a standard upper ontology. In B. Smit and C. Welty (Eds.), *Proceedings of the international conference on formal ontology in information Systems* (pp. 2–9). Association for Computing Machinery.
- Nunberg, G. (2010, December 16). Humanities research with the Google Books corpus. *Language Log*. <https://languagelog.ldc.upenn.edu/nll/?p=2847>
- Pan, H. T., & Zhang, M. F. (2016). Translating for a healthier gaming industry keywords and translation in Macao's gaming discourse. *Translation Spaces*, 5(2), 163–180.
- Pan, H. T., & Zhang, M. F. (2017). Construction and development of Macau's gaming discourse in the post handover era. *Journal of Macau Studies*, 85(2), 73–87. (in Chinese)
- Pease, A. (2011). *Ontology: A practical guide*. Articulate Software Press.
- Pease, A., & Fellbaum, C. (2010). Formal ontology as interlingua: The SUMO and WordNet linking project and global WordNet. In C. R. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, & L. Prevot (Eds.), *Ontology and the lexicon: A natural language processing perspective* (pp. 25–35). Cambridge University Press.
- Potenza, M. N., Balodis, I. M., Derevensky, J., Grant, J. E., Petry, N. M., Verdejo-Garcia, A., & Yip, S. W. (2019). Gambling disorder. *Nature Reviews Disease Primers*, 5, Article 51.
- Rayson, P., & Garside, R. (2000, October). Comparing corpora using frequency profiling. Proceedings of the Workshop on Comparing Corpora, Hong Kong.
- Schützler, O. (2018). Grammaticalisation and information structure: Two perspectives on diachronic changes of notwithstanding in written American English. *English Language and Linguistics*, 22(1), 101–122.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Blackwell.
- Strong, S. (2018). Gamer-generated language and the localisation of massively multiplayer online role-playing games [Ph.D. Thesis]. University College London.
- Teubert, W., & Krishnamurthy, R. (2007). *Corpus linguistics: Critical concepts in linguistics*. Routledge.
- Thompson, W. N. (2009). *The international encyclopedia of gambling*. ABC-CLIO.
- Thompson, W. N. (2015). *Gambling in America: An encyclopedia of history, issues, and society* (2nd ed.). ABC-CLIO.
- Tsai, M. C., Huang, C. R., Chen, K. J., & Ahrens, K. (1998). Towards a representation of verbal semantics—An approach based on near-synonyms. *International Journal of Computational Linguistics and Chinese Language Processing*, 3(1), 61–74.
- Twenge, J. M., VanLandingham, H., & Keith Campbell, W. (2017). The seven words you can never say on television: Increases in the use of swear words in American books, 1950–2008. *SAGE Open*, 7(3), Article 2158244017723689.
- Wang, S., & Huang, C. R. (2017). Word sketch lexicography: New perspectives on lexicographic studies of Chinese near synonyms. *Lingua Sinica*, 3(1), 1–22.
- Wilson, D., & Sperber, D. (2006). Relevance theory. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 607–632). Wiley.
- Wolf, M. J. (2012). *Encyclopedia of video games: The culture, technology, and art of gaming* (Vol. 1). ABC-CLIO.
- Yan, X. (2019). A study of language choices in the linguistic landscape of Macao's heritage and gaming tourism. *Journal of Multilingual and Multicultural Development*, 40(3), 198–217.
- Yoong, D., Koon, T. H., & Min, N. C. (2013). 'This is not gambling but gaming': Methods of promoting a lottery gaming company in a Malaysian daily. *Discourse Studies*, 24(2), 229–247.